

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2004, 21:36:50 ; Search time 6362.59 Seconds

(without alignments)
10782.626 Million cell updates/sec

Title: US-09-938-842A-3729

Perfect score: 1677

Sequence: 1 ggttaagcgtttactatg.....ttctctcagctatattta 1677

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*
1: gb_da.*
2: gb_neg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_da.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rtd.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htg_hum.*
40: em_htg_mus.*
41: em_hgc_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Match | Query Length | DB ID | Description |
|------------|-------|---------|--------------|-------------|----------------------|
| 1 | 1677 | 100.0 | 1677 | AX509034 | AX509034 Sequence |
| 2 | 1677 | 100.0 | 91854 | AC003680 | AC003680 Arabidops |
| 3 | 242.4 | 14.5 | 1483 | AF085354 | AF085354 Arabidops |
| 4 | 102 | 6.1 | 1660 | AY056214 | AY056214 Arabidops |
| 5 | 59.8 | 3.6 | 1141 | AX083744 | AX083744 Sequence |
| 6 | 57.6 | 3.4 | 300695 | AC079431 | AC079431 Mus muscu |
| 7 | 55.6 | 3.3 | 94348 | AC109580 | AC109580 Mus muscu |
| 8 | 54.4 | 3.2 | 1058 | AF015676 | AF015676 Myrmecia pl |
| 9 | 53.4 | 3.2 | 89551 | AP004952 | AP004952 Locust jap |
| 10 | 53.2 | 3.2 | 151143 | AC141740 | AC141740 Apis mell |
| 11 | 53.2 | 3.2 | 249943 | AB014823 | AB014823 Plasmodiu |
| 12 | 52.6 | 3.1 | 56956 | AC117081 | AC117081 Dictyoste |
| 13 | 52.4 | 3.1 | 2000 | AX461144 | AX461144 Sequence |
| 14 | 52.4 | 3.1 | 94038 | AC004401 | AC004401 Arabidops |
| 15 | 51.8 | 3.1 | 34750 | AY217738 | AY217738 Eimeria t |
| 16 | 51.4 | 3.1 | 185747 | EX511261 | EX511261 Dario rer |
| 17 | 51.4 | 3.1 | 349174 | AB063522 | AB063522 Mygilesw |
| 18 | 51 | 3.0 | 65644 | AL590364 | AL590364 Human DNA |
| 19 | 51 | 3.0 | 257109 | AC115577 | AC115577 Dictyoste |
| 20 | 50.8 | 3.0 | 15635 | AB083339 | AB083339 Bombyx mo |
| 21 | 50.8 | 3.0 | 15643 | AF149768 | AF149768 Bombyx mo |
| 22 | 50.8 | 3.0 | 15656 | AB070264 | AB070264 Bombyx mo |
| 23 | 50.8 | 3.0 | 15664 | AF048187 | AF048187 Bombyx mo |
| 24 | 50.8 | 3.0 | 15928 | AB070263 | AB070263 Bombyx ma |
| 25 | 50.6 | 3.0 | 8056 | AX599046 | AX599046 Sequence |
| 26 | 50.6 | 3.0 | 183648 | AC117076 | AC117076 Dictyoste |
| 27 | 50.4 | 3.0 | 7768 | SVU28974 | SVU28974 Spiroplasma |
| 28 | 50.4 | 3.0 | 340552 | PR929354 | PR929354 Plasmodiu |
| 29 | 50 | 3.0 | 810 | CNS06EWO | AL395302 T7 end of |
| 30 | 50 | 3.0 | 969 | DNA132902 | AL122902 Drosophila |
| 31 | 50 | 3.0 | 335050 | PF929356 | AL929356 Plasmodiu |
| 32 | 49.8 | 3.0 | 244674 | AL831726 | AL831726 Dario rer |
| 33 | 49.4 | 2.9 | 110000 | PFMAL13_21 | Continuation (22 o |
| 34 | 49.4 | 2.9 | 110000 | PFMAL13FP_2 | Continuation (3 of |
| 35 | 49.4 | 2.9 | 193187 | AC116920 | AC116920 Dictyoste |
| 36 | 49.4 | 2.9 | 333321 | AC116986 | AC116986 Dictyoste |
| 37 | 49.2 | 2.9 | 49843 | AF288092 | AF288092 Naegleria |
| 38 | 49.2 | 2.9 | 160000 | AB090308 | AB090308 Bombyx mo |
| 39 | 49.2 | 2.9 | 179484 | AC010595 | AC010595 Homo sapi |
| 40 | 49.2 | 2.9 | 349980 | AX344569 | AX344569 Sequence |
| 41 | 49 | 2.9 | 369 | AX185123 | AX185123 Sequence |
| 42 | 49 | 2.9 | 64394 | AL928596 | AL928596 Human DNA |
| 43 | 49 | 2.9 | 110000 | PFMAL8P1_11 | Continuation (12 o |
| 44 | 49 | 2.9 | 349980 | AX344558 | AX344558 Sequence |
| 45 | 48.8 | 2.9 | 653 | AF140509 | AF140509 Apis cera |

ALIGNMENTS

RESULT 1
AX509034 1677 bp DNA linear PAT 27-SEP-2002
LOCUS Sequence 3729 from Patent WO0216655.
DEFINITION AX509034
ACCESSION AX509034
VERSION AX509034.1 GI:23390271
KEYWORDS
SOURCE
ORGANISM Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosoids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
AUTHORS Harper, J.F., Kreps, J., Wang, X. and Zhu, T.
TITLE Stress-regulated genes of plants, transgenic plants containing

same, and methods of use
 Patent: WO 0216655-A 3729 28-FEB-2002;
 The Scripps Research Institute (US) ; Syngenta Participations AG
 (CH)

FEATURES
 source Location/Qualifiers

1. 1677
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /db_xref="taxon:3702"

BASE COUNT 508 a 292 c 335 g 542 t

Query Match 100.0%; Score 1677; DB 6; Length 1677;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1677; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GGTAAAGCGTTTACTTATGTTATATGCAACGGAAGATATTGCCATGTTGGAATGC 60
 1 GGTAAAGCGTTTACTTATGTTATATGCAACGGAAGATATTGCCATGTTGGAATGC 60
 61 TTTTCAGATGATCAAGGCTCCCTACAGATTCTTGAAGGATGGTTTCAGGCTTTGTA 120
 61 TTTTCAGATGATCAAGGCTCCCTACAGATTCTTGAAGGATGGTTTCAGGCTTTGTA 120
 121 GAAATTTGTTTATTGCAACAGTATAGACATTAACCATAGACAGATGTAATGAAAGA 180
 121 GAAATTTGTTTATTGCAACAGTATAGACATTAACCATAGACAGATGTAATGAAAGA 180
 181 TAAAGTTCTTATGCTTAAAGAAATGGAACCGATACGAATTAACAGATCATTAAGAT 240
 181 TAAAGTTCTTATGCTTAAAGAAATGGAACCGATACGAATTAACAGATCATTAAGAT 240
 241 TAAAGTTCTTATGCTTAAAGAAATGGAACCGATACGAATTAACAGATCATTAAGAT 300
 241 TAAAGTTCTTATGCTTAAAGAAATGGAACCGATACGAATTAACAGATCATTAAGAT 300
 301 AAAAACAATCGGAATCCAAACCTCAATTACCAATACGACCCAAATTTATGATGCTGCG 360
 301 AAAAACAATCGGAATCCAAACCTCAATTACCAATACGACCCAAATTTATGATGCTGCG 360
 361 TAAATGATGTAATGCTATGATGATGAGCAAAAGTGGTGGCTGCAAAATTAACAATATC 420
 361 TAAATGATGTAATGCTATGATGATGAGCAAAAGTGGTGGCTGCAAAATTAACAATATC 420
 421 TAAATGATGTAATGCTATGATGATGAGCAAAAGTGGTGGCTGCAAAATTAACAATATC 480
 421 TAAATGATGTAATGCTATGATGATGAGCAAAAGTGGTGGCTGCAAAATTAACAATATC 480
 481 AAAGCTAAAGGTTACTACCGGTTGACCGGTTTAAATTTGGGTTAAATTCCTAATCC 540
 481 AAAGCTAAAGGTTACTACCGGTTGACCGGTTTAAATTTGGGTTAAATTCCTAATCC 540
 541 GGAATCGGTTTGTGTTTATCTCAAGGCAAGTTATCGCAATATTTTGAATTTTGAAGT 600
 541 GGAATCGGTTTGTGTTTATCTCAAGGCAAGTTATCGCAATATTTTGAATTTTGAAGT 600
 601 GGTAGGGAATGATGAGGCTGCAATATGTTGGGCTTACGCTTCAACATGTTGGAAG 660
 601 GGTAGGGAATGATGAGGCTGCAATATGTTGGGCTTACGCTTCAACATGTTGGAAG 660
 661 AGAGTAGGATCCAGCTCAGGCGCCACATTCATTTGTTTGAAGCTTTCTTTCTGG 720
 661 AGAGTAGGATCCAGCTCAGGCGCCACATTCATTTGTTTGAAGCTTTCTTTCTGG 720
 721 TGTCTTACGCTCCCTCTTTCTGCTGCTGCTATGTAACATGATAGTGTGTTTGA 780
 721 TGTCTTACGCTCCCTCTTTCTGCTGCTGCTATGTAACATGATAGTGTGTTTGA 780
 781 CCCGAAACAAGTACCAAGATCAAAATTAAGTTAATCGGTTACATTAAGTCCGTCG 840
 781 CCCGAAACAAGTACCAAGATCAAAATTAAGTTAATCGGTTACATTAAGTCCGTCG 840
 841 AACTTACAATCAATTCGATTAATCTTGAATCTGATTTCTAGTTGGTTGATGTTAATAT 900

841 AACTTACAATCAATTCGATTAATCTTGAATCTGATTTCTAGTTGGTTGATGTTAATAT 900
 901 CCGATTTGTAAGTACCAAGTACCAAGTACCAAGTACCAAGTACCAAGTACCAAGTAC 960
 901 CCGATTTGTAAGTACCAAGTACCAAGTACCAAGTACCAAGTACCAAGTACCAAGTAC 960
 961 AAAGACGAGATGAGGATTTTGAATCTTGAAGTACCAAGTACCAAGTACCAAGTAC 1020
 961 AAAGACGAGATGAGGATTTTGAATCTTGAAGTACCAAGTACCAAGTACCAAGTAC 1020
 1021 ATTACCTTTTGGTGGGCGAGCTTGTGACCTTATGATGATGATGATGATGATGATG 1080
 1021 ATTACCTTTTGGTGGGCGAGCTTGTGACCTTATGATGATGATGATGATGATGATG 1080
 1081 GGGTTACAGCTTTTCCATTAATTAAGTAAATCTTTTGGCTTACCAATTAATTAAT 1140
 1081 GGGTTACAGCTTTTCCATTAATTAAGTAAATCTTTTGGCTTACCAATTAATTAAT 1140
 1141 ATTGAATATCTTCCACATTAAGTAAATTTGATGAGGATGAGGATGAGGATGAGGAT 1200
 1141 ATTGAATATCTTCCACATTAAGTAAATTTGATGAGGATGAGGATGAGGATGAGGAT 1200
 1201 AAAGTAGATTTTCAATTTGGAGTATCTAGTATCTAGTATCTAGTATCTAGTATCT 1260
 1201 AAAGTAGATTTTCAATTTGGAGTATCTAGTATCTAGTATCTAGTATCTAGTATCT 1260
 1261 TTTGATTTTGGATTTTGAAGCTTTTCTAGGTTTAAACATTAATTAATTAATTAAT 1320
 1261 TTTGATTTTGGATTTTGAAGCTTTTCTAGGTTTAAACATTAATTAATTAATTAAT 1320
 1321 AAAAAGAAACATTTTGTGAAAGAAATTAAGTTTACGACCCCAATTTGACATGAG 1380
 1321 AAAAAGAAACATTTTGTGAAAGAAATTAAGTTTACGACCCCAATTTGACATGAG 1380
 1381 TCCCATTAATTAATCTGATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 1440
 1381 TCCCATTAATTAATCTGATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 1440
 1441 TCGGAATGATTTTAAAGTATCTAGTATCTAGTATCTAGTATCTAGTATCTAGTAT 1500
 1441 TCGGAATGATTTTAAAGTATCTAGTATCTAGTATCTAGTATCTAGTATCTAGTAT 1500
 1501 AAAGGTTGAATTAATTTCTCACTGTTTCTGCTGCTTAAATTAATTAATTAATTA 1560
 1501 AAAGGTTGAATTAATTTCTCACTGTTTCTGCTGCTTAAATTAATTAATTAATTA 1560
 1561 AATTTAGTATCTTCTGAGAGTAAAGGTTAAAGGTTAAAGGTTAAAGGTTAAAGGTT 1620
 1561 AATTTAGTATCTTCTGAGAGTAAAGGTTAAAGGTTAAAGGTTAAAGGTTAAAGGTT 1620
 1621 GCTTTTAAACATCTGCTGATCTGATCTGACCGGCTGCTTCTGACGATATATTTTA 1677
 1621 GCTTTTAAACATCTGCTGATCTGATCTGACCGGCTGCTTCTGACGATATATTTTA 1677

RESULT 2

AC003680/c

91854 bp DNA linear JUN 11-MAR-2002

LOCUS

DEFINITION

Arabidopsis thaliana chromosome 2 BAC F17K2 genomic sequence,

complete sequence.

ACCESSION

AC003680

VERSION

AC003680.3

GI:20197048

KEYWORDS

HTG

SOURCE

ORGANISM

Arabidopsis thaliana (thale cress)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis

1 (bases 1 to 91854)

REFERENCE

AUTHORS

Lin X., Kaul S., Town C.D., Beutis M.-I., Greasy T.H., Haas B.J.,
 Mu,D., Walch R., Rensing C.M., Koo H., Fujii C.Y., Uteback T.R.,
 Barnstead M.E., Bowman C.L., White O., Nierman W.C. and Fraser C.M.

TITLE Arabidopsis thaliana chromosome 2 BAC F17K2 genomic sequence
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 91854)
 AUTHORS Lin.X.
 TITLE Direct Submission
 JOURNAL Submitted (09-YAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
 3 (bases 1 to 91854)
 REFERENCE Town, C.D. and Kaul, S.
 AUTHORS Direct Submission
 TITLE Submitted (27-FEB-2002) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA, cdtonw@igr.org
 JOURNAL On Apr 18, 2002 this sequence version replaced gi:6598396.
 COMMENT Address all correspondence to: at@tigr.org

BAC clone F17K2 is from Arabidopsis thaliana chromosome 2 and is near the molecular marker (s) FUS.
 The orientation of the sequence is from SP6 to T7 end of the BAC clone.
 Genes were identified by a combination of several methods: Gene prediction programs including Genscan* (Chris Burge, <http://CCR-081.mit.edu/GENSCAN.html>), GenemarkHMM (Mark Borodovsky, <http://genemark.biology.gatech.edu/Genemark/>), GlimmerA (a variant of GlimmerM, see Mihaela Pertea, <http://www.tigr.org/scftlab/glimmerm.htm>), and GeneSplicer (Mihaela Pertea and Steven Salzberg, contact mpertea@tigr.org), searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by repeatmasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

FEATURES
 source
 1..31954
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="2"
 /map="C1C02E07"
 /clone="F17K2"
 complement(1..1281)
 /note="overlap with BAC clone F4L23 (AC002387:1..1281)."
 57..747
 /gene="At2g45480"
 /note="synonym: F17K2.27"
 join(<57..242,335..5747)
 /gene="At2g45480"
 join(57..242,335..747)
 /gene="At2g45480"
 /note="Protein sequence is in conflict with the conceptual translation"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAF18607.2"
 /db_xref="GI:20197053"
 /translation="DESKPSVQKFFPEVSDKCLEAKFSSNRKNDIARSREWNKNNV KGLFHGFHPSDVLQGRGCPRLQGTDPNEGRRCRTDGGKWRCSKDVLSGQKYCD KMHKRGKKKHVPDITNSHENGAFSLTVAVRSVVPCKDQDKHVSVMGTLPLR VSDKSTSCSDTRITITLALGEDDDEYLSLSPGV"
 complement(1002..2388)
 /gene="At2g45490"
 /note="synonym: F17K2.2; contains a protein kinase domain profile (pDOC00100)"
 complement(join(1002..1112,1208..1280,1366..1580,1655..1750,1886..2128,2260..22388))

CDS
 /gene="At2g45490"
 complement(join(1002..1112,1208..1280,1366..1580,1655..1750,1886..2128,2260..2388))
 /gene="At2g45490"
 /codon_start=1
 /product="putative protein kinase"
 /protein_id="AAC06151.1"
 /db_xref="GI:2979542"
 /translation="MSKKSTESDAGNTEKQWGLADFEIGRPLGKGGKFGVYLAREAKS KTVALKVIFKEQIKYKHQIRREMEIQTSLRHPNLRLEFGWHDNERIELLEYA HGGEYLVKQNGHTEQAAVTLASLOALAYCHGKCVIHRDIXPNLLIDHEGRLLK IADFGSVSSNKRKTMCTGLDYLAPEMVENRDHCVADNWTGLICVFEFLVGNPFE ASQKQTFKRLIKIDLSPLFTFNVEEAKNLISQLLVKDPKSLSEIKIMQHPWLVKN ADFKGCASIDI"
 complement(3414..5980)
 /gene="At2g45500"
 /note="synonym: F17K2.3; predicted by genscan"
 complement(join(<3414..3616,3639..3825,3916..3965,4092..4215,4582..4712,4769..4864,4948..5188,5352..5374,5506..5585,5670..5980))
 /gene="At2g45500"
 complement(join(3414..3616,3639..3825,3916..3965,4092..4215,4582..4712,4769..4864,4948..5188,5352..5374,5506..5585,5670..5980))
 /gene="At2g45500"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AAC06152.1"
 /db_xref="GI:2979543"
 /translation="MSFLRGIIDFSSILNEBKDPVSSSTSTSESMNGIDGVPVT NRIAYLKGYPDLAKEIAEKVRAEWEGLHDALLHYRNAQRINNEATSTSPSYIS SEKPEVRSYREKISNMQVSRQALGVGSENKRTVAVPSSASVSTASRYKRTL SOKTPVARGVATPKNPKDAASPKVKGSGNVYDDKLVEINTIVDRSPVKWDDV AGLNGAKALLEMLVPKRRDLFGLRRPARVTSLLGLLFGPGNGKTMKAKAVAS EQATFNVSASSLTSKIDSIMSTRSENSEARSLRLESEFLIQDVTNSPDDDLVII IGATNPQLCDLAVRLRVKIVFLPDPNSNVRKLLFKLKQCPHSLGDDGDKIVKE TEGKLVLCIKKHRTISQVTDKRYSGSDIQAACEAAMPPIRELGANILTIQANKVLN FSVQINVBVLSLKLINRCVHVTKISTI"
 6429..8288
 /gene="At2g45510"
 /note="synonym: F17K2.4"
 join(<6429..6950,7026..7310,7386..7724,7806..8006,8100..8288)
 /gene="At2g45510"
 join(6429..6950,7026..7310,7386..7724,7806..8006,8100..8288)
 /gene="At2g45510"
 /codon_start=1
 /product="putative cytochrome P450"
 /protein_id="AAC06153.1"
 /db_xref="GI:2979544"
 /translation="MEILTSIATVATTIFVLCTIYIMIRIFTGKSRNDRKVPVH ATVDELPHSDLYDYETIEIAKEKTVRFSPQSEILTADPRNVHILKTRFDNYSK GHSSRENADLLGHGIFAVDGEKMQKQLSPSESTVLRDPSCSVRRNASKLVGF VSEFALSGRAFQADQLLMKTLDSIFKVGVELKCLDGSEGEFWEAFDEGNVAT SRFDPLWLKGFNFNIGSKSLSIATIDKFVYSLITTKRELAKQNTVVRDIL SRFLVSEKDPENMDKYLRIILNFMAGKDTAALLSWFLMKLCKPLVQERTVQE IRDVPFSHKTTDVNGFVSEINEEALDENHYLHAALSETRLRYPVPVDMRCAENDV LDPGHVSGKDNLYIAYAMGMTYIWGDAEEFKPERWLKGLFQPSPPKFI SFHA GPRICLGKDFAYRQMKIVSMALLHFRFPKADENSKVYKMLTLHVDGLHLCAIPR TSI"
 complement(8412..9632)
 /gene="At2g45520"
 /note="synonym: F17K2.5; supported by full length cDNA: Ceres.8220"
 complement(join(<8412..8678,8768..8903,8987..9120,9289..9356,9459..9632))
 /gene="At2g45520"
 complement(join(8553..8678,8768..8903,8987..9120,9289..9356,9459..9612))
 /gene="At2g45520"
 /codon_start=1
 /product="expressed protein"

```
/protein_id="AAC06155.1"
/db_xref="GI:2979546"
/translation="MGKGKKREKNYLAHGGPAPRLPPPPDRSKDDVPSTLIRILNN
YTSPPHDSKQVVEKKEKLVKAEVVPVATESDDSDSVKVKKKKKRQMDLRPEN
ELABIDRSKRKKKKYWEAKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
VVPKARSTLSASQERLRLQALDAYRSRKGWTARPVPFPAVMQ"
9408..9427
/rpt_family="(CA)n"
complement(9578..9632)
/rpt_family="(GA)n"
9822..10798
/gene="At2g45530"
/notes="synonym: F17K2.6; supported by full length cDNA:
Ceres:35578"
9822..10798
/gene="At2g45530"
9948..10670
/gene="At2g45530"
/codon_start=1
/product="expressed protein"
/protein_id="AAC06154.1"
/db_xref="GI:2979545"
/translation="MQSSSSGGDDLEKQQQQQKKSPKQSESANESNHLTIIVCN
GDSERELQVQIPPEKVSUSRNGSHQCRVLDQKEVLIELGCCRGGLAKAHR
CIDAFRTKGSNCEICQVAVNTPPETQPTTNVWMDPSYRQEEERECFSLPW
VAFSILIGMLDVLISITLGVGALPVNIIIGVIVLGLTALRLTLFCYEWSLRA
VHRVQRAEANNFSNIAYPAL"
9988..10007
/rpt_family="(CAG)n"
complement(10833..125573)
/gene="At2g45540"
/notes="synonym: F17K2.7"
complement(join(<10833..11272,11438..11560,11636..12378,
12454..12937,13023..13119,13254..13374,13464..13548,
13645..13794,13957..14073,14178..14264,14350..14418,
14520..14576,14730..14813,14928..14995,15396..15538,
15768..16244,16322..16465,16545..17093,17237..17431,
17612..17696,18015..18613,18944..19249,19633..19714,
19972..20048,20151..20264,20429..20533,20796..20930,
21000..21007))
Query Match 100.0%; Score 1677; DB 8; Length 91854;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1677; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGTAAAGCGTTTACTTATGTTTATATGCAACGAGAGATATGCGATGTTGGATGC 60
DB 76808 GGTAAAGCGTTTACTTATGTTTATATGCAACGAGAGATATGCGATGTTGGATGC 76749
QY 61 TTTTTCAGATCATCAAGGCTCCTACAGATTTCTTAGGGAATGTTTCAGGCTTTTGTTA 120
DB 76748 TTTTTCAGATCATCAAGGCTCCTACAGATTTCTTAGGGAATGTTTCAGGCTTTTGTTA 76689
QY 121 GAAATGTTTATGCAACAGGTAGAGAACATAACCATAGACAGATGATCTGAAGAGA 180
DB 76688 GAAATGTTTATGTTCAACAGGTAGAGAACATAACCATAGACAGATGATCTGAAGAGA 76629
QY 181 TAAGCTTCTATGCTCAAGAAATGACCGATACGATACGATCAATTAAGAT 240
DB 76628 TAAGCTTCTATGCTCAAGAAATGACCGATACGATACGATCAATTAAGAT 76569
QY 241 TAAATGTTTGTGAAGAAATACTACATTTATGTTGAATGTTGTTAGTGAAGAT 300
DB 76568 TAAATGTTTGTGAAGAAATACTACATTTATGTTGAATGTTGTTAGTGAAGAT 76509
QY 301 AAAAATCGGAATCCAAATCCTCAATTTACCAATCAGCCCAAAATTTATGTCGTGGC 360
DB 76508 AAAAATCGGAATCCAAATCCTCAATTTACCAATCAGCCCAAAATTTATGTCGTGGC 76449
QY 361 TAATGAATGTTATGCTCATGAGGAAAAGTTGGTGGCTGCGAATTAACACATATC 420
DB 76448 TAATGAATGTTATGCTCATGAGGAAAAGTTGGTGGCTGCGAATTAACACATATC 76389
QY 421 CCTCTGGTGGACCCGAAATCTGTAATCGGAAAGGTGGAAACCCACTTGGTTTAACTTTT 480
```

```
DB 76388 CCTCTGTGTGGACCCGAAATCTGTAATCGGAAAGGTGGAAACCCACTTGGTTAACTTTT 76329
QY 481 AAGCCATAAAGGTTACTACCGGTTTGAACCGGTTTATAATTTGGTGTAAATCTCTAAATCC 540
DB 76328 AAGCCATAAAGGTTACTACCGGTTTGAACCGGTTTATAATTTGGTGTAAATCTCTAAATCC 76269
QY 541 GGATCCGTTTGTGTGTTAATCTAAGGCCACGTTATCGCCAATATTTTGTAGTG 600
DB 76268 GGATCCGTTTGTGTGTTAATCTAAGGCCACGTTATCGCCAATATTTTGTAGTG 76209
QY 601 GTTAGGGAAATGTCGGGTCGAAATAGTTGGGCTAGCCCTCAACATGTCGTGGAACTCAAG 660
DB 76208 GTTAGGGAAATGTCGGGTCGAAATAGTTGGGCTAGCCCTCAACATGTCGTGGAACTCAAG 76149
QY 661 AGATAGGGTCCAGCTCAGGCCCAATTCACATTTTCGTTTGTAGCCTTTCTTTCTGG 720
DB 76148 AGATAGGGTCCAGCTCAGGCCCAATTCACATTTTCGTTTGTAGCCTTTCTTTCTGG 76089
QY 721 TGCTTAGGTCCTCTTCTGTCGGTCGTATGTAAGTAGCATAGTAGTGGTTCAAA 780
DB 76088 TGCTTAGGTCCTCTTCTGTCGGTCGTATGTAAGTAGCATAGTAGTGGTTCAAA 76029
QY 781 CCGGAAACAAGTACCAACGAATCAAAATAAGTTTGAATCGGTTACATCTAGTACCGTCG 840
DB 76028 CCGGAAACAAGTACCAACGAATCAAAATAAGTTTGAATCGGTTACATCTAGTACCGTCG 75969
QY 841 AACTTACATCATTTCAATTTCTGATCTGATTTCTAGTTCGTTTGTAGTTTAAAT 900
DB 75968 AACTTACATCATTTCAATTTCTGATCTGATTTCTAGTTCGTTTGTAGTTTAAAT 75909
QY 901 CCGGATTTGTAACAGTACACAAGTACATAAGTATGTCGTATATGTAACCGGTTTAAATC 960
DB 75908 CCGGATTTGTAACAGTACACAAGTACATAAGTATGTCGTATATGTAACCGGTTTAAATC 75849
QY 961 AAAGAACGAGATAGAGAGATTTTGAATCTGGAAGAGGATTAFTCCATAGACATA 1020
DB 75848 AAAGAACGAGATAGAGAGATTTTGAATCTGGAAGAGGATTAFTCCATAGACATA 75789
QY 1021 ATTAGCTTTTGGTGGCGCAGGCTTGTGACCTACATTAATGCGGTCACACCCCAAGTATG 1080
DB 75788 ATTAGCTTTTGGTGGCGCAGGCTTGTGACCTACATTAATGCGGTCACACCCCAAGTATG 75729
QY 1081 GGTTCACGCTTTTCCATAAAATTAAGTAAATCTTTTTCGCTTAAACCAATATAAAAT 1140
DB 75728 GGTTCACGCTTTTCCATAAAATTAAGTAAATCTTTTTCGCTTAAACCAATATAAAAT 75669
QY 1141 ATTGAAATCTTCCACACCAATAGAAAGTTAAATTTGATCAGCGATGGAAATTTTGTAC 1200
DB 75668 ATTGAAATCTTCCACACCAATAGAAAGTTAAATTTGATCAGCGATGGAAATTTTGTAC 75609
QY 1201 AAAGCTAGGTATTTCAATTTGGGAGTGTACTAGTAACTAGTAAGTACTAACCAAGATGAT 1260
DB 75608 AAAGCTAGGTATTTCAATTTGGGAGTGTACTAGTAACTAGTAAGTACTAACCAAGATGAT 75549
QY 1261 TTCTGATTTGATTTGAGGCTTTTCTTAGGTTTAAATAAAGTATATTAATAAAT 1320
DB 75548 TTCTGATTTGATTTGAGGCTTTTCTTAGGTTTAAATAAAGTATATTAATAAAT 75489
QY 1321 AAAAGAAAAACATTTTGTGAAAAAGAGAAATTAAGTTTACTGGAACCCCATTTGACAGATG 1380
DB 75488 AAAAGAAAAACATTTTGTGAAAAAGAGAAATTAAGTTTACTGGAACCCCATTTGACAGATG 75429
QY 1381 TCCCATTAATACTATGATAGAGATAGAGCAATGGAAGGATTTGTTTCACTGGTGGTACAA 1440
DB 75428 TCCCATTAATACTATGATAGAGATAGAGCAATGGAAGGATTTGTTTCACTGGTGGTACAA 75369
QY 1441 TCGGAATGTTTCTTTAAGCTCATCGAACACATCAGACCGGTTGATTTTCCGCGATCAA 1500
DB 75368 TCGGAATGTTTCTTTAAGCTCATCGAACACATCAGACCGGTTGATTTTCCGCGATCAA 75309
QY 1501 AAAGCGTTTGAATACATTTCTCACTTGTTCCTGCTTATATATATATCTCTGACGATC 1560
```



```

/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="2"
/clone="RAFLO7-08-P04 (R10678)"
/note="This clone is in a modified pBluescript vector (FUC-1) as a BamHI/XhoI insert."
ecotype: Columbia
1..1660
/gene="At2g45680"
1'..102
/gene="At2g45680"
103..1173
/gene="At2g45680"
/codon_start=1
/evidence=experimental
/product="putative PCF2 DNA binding protein"
/protein_id="AAU07063.1"
/db_xref="GI:15810351"
/tranlations="MATIOKLEVAGKDOTLRADLTITINGVENVETSPFQVNPVTVS
LESAEPVWPSFMSLAPSSGTPPLKRASTKDRHTKVEGRGRIIMPATCAARIQQL
TRELGHKSDBETIRWLENAEPALIAATGTGTVPALAMSUNGTLKPTITNADSDMGE
NLMKKKRSPENSEYIDISDAVSAAGSLAFIATTTIQPQALASSTVAQQLPQMGY
PMWAIISNMIPGTGFVFIPIQAGSNQQLIAFPAAASPSYVVAAVQQAATWARP
PPLQVWSPGCVSVSDVSGNLSGRATSWAPSSSGVTTCSSSSIATTTTHTLRDFSL
EIVEKQELHQFMSTTTARSNH"
1174..1660
/gene="At2g45680"
464 a 400 c 385 g 411 t
3'UTR
BASE COUNT
ORIGIN

```

| | Query Match | 6.1% | Score 102; | DB 8; | Length 1660; |
|----|-----------------------|--|------------------|-----------|--------------|
| | Best Local Similarity | 100.0%; | Pred. No. 5e-11; | | |
| | Matches 102; | Conservative 0; | Mismatches 0; | Indels 0; | Gaps 0; |
| QY | 1576 | CTTGGACGCTAACGCCGCTTAAACGATCTCTTCCCATTCGTATCCGCTTTTAAACAACCTCT | | | 1635 |
| Db | 1 | CTTGGACGCTAACGCCGCTTAAACGATCTCTTCCCATTCGTATCCGCTTTTAAACAACCTCT | | | 60 |
| QY | 1636 | CGTCGTCATCTCACCGTCGCTTTCTCTCAGCTATATTTTA | | | 1677 |
| Db | 61 | CGTCGTCATCTCACCGTCGCTTTCTCTCAGCTATATTTTA | | | 102 |

| | | | | | | |
|------------|---|-------------|-----|--------|--|-----------------|
| RESULT 5 | | | | | | PAT 28-FEB-2001 |
| AX083744/c | | | | | | |
| LOCUS | AX083744 | 1141 bp | DNA | linear | | |
| DEFINITION | Sequence 22 from Patent WO011061. | | | | | |
| ACCESSION | AX083744 | | | | | |
| VERSION | AX083744.1 | GI:13185472 | | | | |
| KEYWORDS | . | | | | | |
| SOURCE | synthetic construct | | | | | |
| ORGANISM | synthetic construct | | | | | |
| | artificial sequences. | | | | | |
| REFERENCE | 1 | | | | | |
| AUTHORS | Kunst, L. and Clemens, S. | | | | | |
| TITLE | Regulation of embryonic transcription in plants | | | | | |
| JOURNAL | Patent: WO 011061-A 22 15-FEB-2001; | | | | | |
| | UNIVERSITY OF BRITISH COLUMBIA (CA) | | | | | |
| FEATURES | Location/Qualifiers | | | | | |
| SOURCE | 1..1141 | | | | | |

promoter
1...1141
/notes=consensus sequence of A.t., L.a., and B.n. FAEI
Promoters
BASE COUNT 123 a 42 g 112 t 832 others
ORIGIN
Query Match 3.6%; Score 59.8; DB 6; Length 1141;
Best Local Similarity 10.5%; Pred. NO. 0.031;
Matches 85; Conservative 299; Mismatches 422; Indels 3; Gaps 1;

| | | | |
|---|------|--|------|
| QY | 855 | TCGATTACTTTGATCTGATTTCTAGTTCGGTTTGTATGTTTAAATATCCGATGTGTACAAG | 914 |
| Db | 1002 | TNNVWYATTTAAAYAAAKWAWAGNNRMVGAAGNKGWCAAMATWGBWADTATGKWC | 943 |
| QY | 915 | TACAAGTACATAAGATATGCGTATATGTATGTGACCGGTTTAAATCAAAAGGACGACGAT | 974 |
| Db | 942 | NNNNNNWTTDVRMAKAKNNNNNNAYMTACYNRAATNNKMATHWMMKWTGHAHSKRTSH | 883 |
| QY | 975 | AGGAGATTTTGAATCTCGGAAAGAGGATTTTCCATAGACACTAATTAATGCTTTTGGT | 1034 |
| Db | 882 | HTTCTCRTKYNNNNNNNIARIVYWHHAAREWMAWMTFTNNNNNNNNNACRNTRTTWAABW | 823 |
| QY | 1035 | GGCGCAGCCTTGTGACCTACATTAATGGGGTCCAAACCCCAAGTATGGGCTTACAGCTTTT | 1094 |
| Db | 822 | XHSWCNNNNNNNNNNNNNTWCHYTTANABBCYRANNNNAARMARTCNMYHAAVTTTH | 763 |
| QY | 1095 | TCCATAAAATTAAGTAATCTTTTTTTCCTTAAACCAATAAAATTAATGAAATCTT--- | 1152 |
| Db | 762 | TDCYCTWNTTWDMTTMBTTTTTTRNNMTTSTNNNNNNNNNNWACTNNNNNNNNKAYAH | 703 |
| QY | 1153 | -TCCAAACCATAGAAAAGTTAAATTTGATCAGCGATGGAAATTTTGTACAAAAGCTAGTA | 1211 |
| Db | 702 | ATNNWCWNNTDARFTNNNTVMBRRWNTKTRWVSTTRHHYTGATNNNNNNNNNNNN | 643 |
| QY | 1212 | TTTCATTTGGGAGTGTACTAGTAACTAGTAAGTACTAACCCAGAAATGAGTTCTCGATTTG | 1271 |
| Db | 642 | NNNNSCCTCTRMWTRWTKGDMGTVRKVKVRDITCTYVDYWADSWVWYANWMCREDV | 583 |
| QY | 1272 | GATTTTGAAGCTTTTCTTAGGTTTAAAAACCAAGTATATTACTAAACAAATAAAAGAAAAAC | 1331 |
| Db | 582 | TYTRNNNTYCKSYAHSYVWYNSNNAMVRRYSARNWSSMARWTTNNNNWNSGBVMBRWGTM | 523 |
| QY | 1332 | ATTTGTGAAAAGAGAAATAAGTTTACTGGAACCCCATTTGTACAGATGTCCTCCATAATA | 1391 |
| Db | 522 | WHRNNNNNTDTRYVWVWVRRWABTTTTVYDSNCNAKSMWRGNWNRAMKWMAANNIDAGAM | 463 |
| QY | 1392 | TACTGATAGAAGATAGACCAATGGAAAGTGATTGTTGTCAGCTGGGTACAATCGCAATGGTT | 1451 |
| Db | 462 | DHWTYMGNTTWMMRRAWOMNNWMCRRAYCCNNNNNRACVWHKHOMRWTKYMKWKAAC | 403 |
| QY | 1452 | CTTTAAAGTCTATCGAACACATCAGACCGGTTGATTTTCCCGCATCAAAAAGCGTTGAA | 1511 |
| Db | 402 | NNNNBKAMTRVAWMMYSRDTTNTDMMWMTSDWBWHWTVDYTMRAWNNNNNNNNWBRCK | 343 |
| QY | 1512 | TACTATCTCAGTCTGTTTCTCGCTCCATATATATATCTCAGCAGTCCACATTTAGTAA | 1571 |
| Db | 342 | TTSSWMMWDMHTHCTYGNNTWGSAYBVAWMSWAGASNBVTYKWCWRMTYMGKTWNTN | 283 |
| QY | 1572 | TCCTCTGGACGTGAACGCCGTTTAAACGATCTTTCCCATGTTATCGCTTTTAAACAA | 1631 |
| Db | 282 | NNNNKAWYRTKTVACNNRNYDVTAVMTBKENYKYCYABWYBYBNTAGKHBBWRRAB | 223 |
| QY | 1632 | CNCTCGTCTGTCATCTCCACCGTCGGTTTTT | 1660 |
| Db | 222 | HRSWNWVVKRNKTVSWHYHAMRYBKW | 194 |
| RESULT 6 | | | |
| AC079431 | | | |
| LOCUS | | | |
| DEFINITION | | | |
| Mus musculus chromosome 16 clone RP23-1J12, WORKING DRAFT SEQUENCE, | | | |
| 77 unordered pieces. | | | |
| AC079431 | | | |
| ACCESSION | | | |
| VERSION | | | |
| KEYWORDS | | | |
| SOURCE | | | |
| ORGANISM | | | |
| Mus musculus | | | |
| Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | |
| Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | | |
| REFERENCE | | | |
| AUTHORS | | | |
| DOE Joint Genome Institute. | | | |
| TITLE | | | |
| Sequencing of Mouse | | | |

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Unpublished
2 (bases 1 to 300695)
DOE Joint Genome Institute.
Direct Submission
Submitted (01-SEP-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 0
Center Clone Name: RPI-23_1J12

Summary Statistics
Consensus quality: 164936 bases at least Q40
Consensus quality: 197223 bases at least Q30
Consensus quality: 214345 bases at least Q20
Estimated insert size: 257200; agarose-fp estimation
Quality coverage: 3.01 in Q20 bases; agarose-fp estimation
Quality coverage: 2.65 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 77 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
1076: contig of 1076 bp in length
1077
1176: gap of unknown length
1177
2208: contig of 1032 bp in length
2209
2308: gap of unknown length
2309
3527: contig of 1219 bp in length
3528
3627: gap of unknown length
3628
4764: contig of 1137 bp in length
4765
4864: gap of unknown length
4865
5952: contig of 1088 bp in length
5953
6052: gap of unknown length
6053
7182: contig of 1130 bp in length
7183
7282: gap of unknown length
7283
8420: contig of 1138 bp in length
8421
8520: gap of unknown length
8521
9616: contig of 1096 bp in length
9617
9716: gap of unknown length
9717
11229: contig of 1413 bp in length
11230
11229: gap of unknown length
11230
12408: contig of 1179 bp in length
12409
12508: gap of unknown length
12509
13767: contig of 1259 bp in length
13768
13867: gap of unknown length
13868
15044: contig of 1177 bp in length
15045
15144: gap of unknown length
15145
16325: contig of 1181 bp in length
16326
16425: gap of unknown length
16426
17578: contig of 1153 bp in length
17579
17678: gap of unknown length
17679
18445: contig of 1167 bp in length
18446
18445: gap of unknown length
18446
20060: contig of 1115 bp in length
20061
20160: gap of unknown length
20161
21258: contig of 1098 bp in length
21259
21358: gap of unknown length
21359
22390: contig of 1032 bp in length
22391
22490: gap of unknown length
22491
23599: contig of 1109 bp in length
23600
23699: gap of unknown length
23700
24907: contig of 1208 bp in length
24908
25007: gap of unknown length
25008
26121: contig of 1114 bp in length
26221: gap of unknown length

26222
27273: contig of 1052 bp in length
27274
27373: gap of unknown length
27374
28472: contig of 1099 bp in length
28473
28572: gap of unknown length
28573
28583: contig of 1111 bp in length
28584
29833: gap of unknown length
29834
30334: contig of 1051 bp in length
30335
30835: gap of unknown length
30836
32035: contig of 1101 bp in length
32036
32135: gap of unknown length
32136
33334: contig of 1199 bp in length
33335
33434: gap of unknown length
33435
34598: contig of 1164 bp in length
34599
34698: gap of unknown length
34699
35818: contig of 1120 bp in length
35819
35918: gap of unknown length
35919
37082: contig of 1164 bp in length
37083
37182: gap of unknown length
37183
38445: contig of 1453 bp in length
38446
38445: gap of unknown length
38446
39768: contig of 1023 bp in length
39769
39868: gap of unknown length
39869
40973: contig of 1105 bp in length
40974
41073: gap of unknown length
41074
42174: contig of 1101 bp in length
42175
42274: gap of unknown length
42275
43341: contig of 1067 bp in length
43342
43441: gap of unknown length
43442
44532: contig of 1091 bp in length
44533
44632: gap of unknown length
44633
46068: contig of 1436 bp in length
46069
46168: gap of unknown length
46169
47421: contig of 1253 bp in length
47422
47521: gap of unknown length
47522
48629: contig of 1108 bp in length
48630
48729: gap of unknown length
48730
49886: contig of 1257 bp in length
49887
50086: gap of unknown length
50087
51394: contig of 1308 bp in length
51395
51494: gap of unknown length
51495
53158: contig of 1664 bp in length
53159
53258: gap of unknown length
53259
54802: contig of 1544 bp in length
54803
54902: gap of unknown length
54903
56009: contig of 1107 bp in length
56010
56109: gap of unknown length
56110
57452: contig of 1343 bp in length
57453
57552: gap of unknown length
57553
58667: contig of 1115 bp in length
58668
58767: gap of unknown length
58768
59224: contig of 1157 bp in length
59225
59925: gap of unknown length
59926
60024: gap of unknown length
60025
61159: contig of 1135 bp in length
61160
61259: gap of unknown length
61260
62764: contig of 1505 bp in length
62765
62864: gap of unknown length
62865
64293: contig of 1429 bp in length
64294
64393: gap of unknown length
64394
65793: contig of 1400 bp in length
65794
65893: gap of unknown length
65894
67168: contig of 1275 bp in length
67169
67268: gap of unknown length
67269
69245: contig of 1977 bp in length
69246
69345: gap of unknown length
69346
70850: contig of 1505 bp in length
70851
70850: gap of unknown length
70851
73220: contig of 2270 bp in length
73221
73220: gap of unknown length
73221
76908: contig of 3588 bp in length
76909
77008: gap of unknown length
77009
78773: contig of 1765 bp in length
78774
78873: gap of unknown length
78874
81789: contig of 2916 bp in length


```

Db      88200 TATTATTAAATGCAATAAATAAATTAAATTT 88233

RESULT 8
LOCUS   MP15676 1058 bp DNA linear INV 09-FEB-1996
DEFINITION Myrmecia pilosula HI87-128 mitochondrion cytochrome b gene, partial cds.
ACCESSION U15676
VERSION   U15676.1 GI:576754
KEYWORDS  mitochondrion Myrmecia pilosula (Australian jumper ant)
SOURCE    Myrmecia pilosula
ORGANISM  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Formicidae; Myrmecinae; Myrmecini; Myrmecia.
REFERENCE 1 (bases 1 to 1058)
AUTHORS   Crozier,R.H., Dobric,N., Imai,H.T., Graur,D., Cornuet,J.M. and Taylor,R.W.
TITLE     Mitochondrial-DNA sequence evidence on the phylogeny of Australian Jack-jumper ants of the Myrmecia pilosula complex
JOURNAL   Mol. Phylogenet. Evol. 4 (1), 20-30 (1995)
MEDLINE   95346030
PUBMED    7620633
REFERENCE 2 (bases 1 to 1058)
AUTHORS   Crozier,R.H.
DIRECT SUBMISSION
TITLE     Submitted (08-OCT-1994) Ross H. Crozier, School of Genetics, La Trobe University, Bundoora, Victoria 3083, Australia
JOURNAL
FEATURES
         Location/Qualifiers
         1..1058
            /organism="Myrmecia pilosula"
            /organella="mitochondrion"
            /mol_type="genomic DNA"
            /strain="HI87-128"
            /db_xref="taxon:13618"
            /issue_type="head and thorax muscle"
            /dev_stage="adult worker"
            <2..733
            /codon_start=1
            /transl_table=5
            /product="cytochrome b"
            /protein_id="AA86970.1"
            /db_xref="GI:576755"
            /translation="FMGATVTLNVLSTPFLGSDITQWLWGYSVSNATLGRFSIHF
            ILPMLIMVIMVHLPLHTGNSPLGINSNLKIPFNLYSTKDLGLFTTLTLLFTL
            LLMTPLSDPDNPLPANPLVTPHIOPEWLEFAYSLRSIPNKLGVIALPSSIC
            ILVFLPLNSNNSSTFPLSQLLINLFWMTFMTLTLGSQLSVIEHPYIIMAQFFSMLY
            FNFILLKFSYKXWDYFIYNKFNN"
            /misc_feature 734..1058
            /notes="intergenic space"
BASE COUNT 398 a 132 c 61 g 467 t
ORIGIN
Query Match 3.2%; Score 54.4; DB 3; Length 1058;
Best Local Similarity 48.1%; Pred. No. 0.42;
Matches 154; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

2Y      1092 TTTTCCATAAAATTAAGTAAATCTTTTTCCTAACCAATAAAATTTATTGAAATCT 1151
      |||
2b      739 TTTTATTAAATAATTTATTTAATAATATTTATTTAATAATATTTATTTAATAATA 798
      |||
2Y      1152 TTCCAACCATAGAAAGTTAAATTTGATCAGCGATGGAATTTTGTACAAAGCTAGGTA 1211
      |||
2b      799 TTTTATTGATTAAAAATTAATTTAAATTTAAATTTAAATTTTATAAAATTTAAATTT 858
      |||
2Y      1212 TTTTCATTTCGGAGTACTAGTACTAGTACTAGTACTAGTACTAGTACTAGTACTAGT 1271
      |||
2b      859 ATTATAATTTTATTTTTATTTAATAATAATTTATTTATTTATTTATTTATTTCTTTA 918
      |||
2Y      1272 GATTTTGAAGCTTTTCTTAGTTAAAAAACAAGTATATTACTTAAACAATAAAGAAAAAC 1331
      |||
2b      919 TAAATTTATATATAATTTATATATAACAATAAAATTTATTATATTATATAAATAATA 978
      |||

```

```

QY      1332 ATTTTGTGAAAGAGAAATAAAGTTTACTGGAACCCCATTTGTACAGATGGTCCCATATAA 1391
      |||
Db      979 ATTATATTAAATTTCTATCTAACTAACTAACTAAATTTTATTTTAAATAAATAATCA 1038
      |||
QY      1392 TACTGATAGAAGATAGAGCA 1411
      |||
Db      1039 AAAATATTTAAGTTAATGAA 1058
      |||

RESULT 9
LOCUS   AP004952/c 89551 bp DNA linear PLN 19-JUL-2002
DEFINITION Lotus japonicus genomic DNA, chromosome 1, clone: LJ19C08, TM0125, complete sequence.
ACCESSION AP004952
VERSION   AP004952.1 GI:21907970
KEYWORDS  HTG.
SOURCE    Lotus japonicus
ORGANISM  Lotus japonicus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotaeae; Lotus.
REFERENCE 1
AUTHORS   Kaneko,T., Nakamura,Y., Asamizu,E., Kato,T., Sato,S. and Tabata,S.
TITLE     Structural Analysis of a Lotus japonicus Genome. I. Sequence Features and Mapping of Sixty-six TAC clones which cover the 6.7 Mb Regions of the Genome
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 89551)
AUTHORS   Nakamura,Y.
DIRECT SUBMISSION
TITLE     Submitted (26-MAR-2002) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: ynk@kazusa.or.jp, URL: http://www.kazusa.or.jp, Tel: 81-438-52-3935, Fax: 81-438-52-3934)
JOURNAL
FEATURES
         Location/Qualifiers
         1..89551
            /organism="Lotus japonicus"
            /mol_type="genomic DNA"
            /db_xref="taxon:34305"
            /chromosome="1"
            /clone_lib="J1T library"
            /clone="LJ19C08"
            /notes="TAC clone: TM0125"
BASE COUNT 27663 a 17797 c 17022 g 27069 t
ORIGIN
Query Match 3.2%; Score 53.4; DB 8; Length 89551;
Best Local Similarity 45.6%; Pred. No. 0.34;
Matches 224; Conservative 0; Mismatches 266; Indels 1; Gaps 1;

QY      1055 ATTAATGGGTCCCAACCCCAAGTATGGGCTTTCAGCTTTTCCATAAAATTAAGTAAT 1114
      |||
Db      73682 ACTAATTTGGAATGACTGAAAAGTAACTAAATAATTTATCCTTTAAATAAATAGTTT 73623
      |||
QY      1115 CTTTTTCCTCAACCAATAAAATTTATGAAATCTTTCCACCACATAGAAAGTTAAAT 1174
      |||
Db      73622 CATATTTTAATAATTTTATTATTATTGAAATTAATCAATAATGATACAAAGATACAA 73563
      |||
QY      1175 TTGATCAGCGATGGAATTTTGTACAAAGCTAGGTATTTCATTTTGGGAGTGTACTAGTA 1234
      |||
Db      73562 ATGATAATTTATTATTTTACATAAATGTAATAATTAATTCCTTTAAATATATTTACTGTA 73503
      |||
QY      1235 ACTAGTAAGTACTAACCGAAGTGTCTGATTTTGGATTTTGAAGCTTTTCTTAGGTT 1294
      |||
Db      73502 AATAGACGAAATTAATAATTAATTAATAGATTTTGTATATAAAGTATT -TTTTFG 73444
      |||
QY      1295 AAAAAACAAGTATATTACTTAAACAATAAAGAAAAACATTTTGTGAAAGAGAAATAAAG 1354
      |||
Db      73443 AAAATGTATAAATAATTTACTTCTATAAATAAAGTATTATTTTTAAATAATAATAATAA 73384
      |||

```


YTNISENLYRKISITYETSSRSDSDNVLLENINNVLVKTVSGLDVTLNIRISRDK
 IBEIKFYPIIHNNVQLANDYIYSDEVINLLKQKONIKCLDKQMLFIIEKNVDYFK
 NMLKPGFYIPCNKFLTKYKQNDNERCLTDTVDNDKSNSTTNSYILSSPNMNTNT
 NYSNHSNIIARNTSSNKNQSPIGFINERNSSPNFGIENESIKWKKQDQNMDDPVYLI
 MEYNLESHVILADTSSNKNPNIESMDNRYSNPNFIEHLIHNDDTOJNVDYDI
 NNMGGKKNLNSNTSDGTFQSISDFYKNEKKILLCOEONNTYIKQCAFINSNETV
 LTNKNNHSHKQSLISLDFDELSALTINDSSLNKAIFYENFNSEKQIVEDIQNT
 NYXDIHMEHNTYQNDERKMDDEYGDNYNIPRSKQILKKCKELLISCIYSNEKENYK
 DQNMILNEDNTKVNLIYDIGNISPDENIGINNHNNTNINLVITYEANKSYGDNNE
 RLNFIRKQJILLNKYKVKLTKTQSTTLKCLNLLYKVKATDTQVFLPYCHYKMODS
 IEITHDKKNSQFVPLNLTIIKKCKENYRQDITKNLHDKQILINKKKVEPKVC
 LKTYKNGKQFDOGLLEWLVNLICNATYRHHNLSKNLIOLYDSFYVKEHLIIVYEM
 QSDLYNYPFKAGKLTGLQQLIKLKNLEGLAYIHSKNLHCDLKPEBINIMWKKK
 NHEKGYKNQWNGVNIYNDIEPLHNSNLSNLEKNNIAYPFSFDTFTLENKA
 QYDNNEKSNLYDSNKSYPNKNMIDNNLYCNINIKNNNNNNNNNNNNNNPHN
 INYNTQPKDKIKIIPDNSCIYESDKLEMTIQTRYESPVLJQNYDKRIDITWSLGC
 ILIPEFTKILIPDYQNIYRFYIYSIVSYGPPFYMINNCRIPIHIFKHGLIILKCTT
 NNWNEIEKEQLOEDDEBEIVFNSNDPFRLNKQNTILTKOLLKKNPNTSTPKRNN
 DNNNEIYDYVCYSDNLLKNKQFISDTLFPVDFLSLLQIDPSKCSNAMEALKHPWLQPN
 LYKGL

1089..1149
 /rpt_type=tandem
 4186..4225
 /rpt_type=tandem
 4358..4387
 /rpt_type=tandem
 4543..4577
 /rpt_type=tandem
 5071..5152
 /rpt_type=tandem
 5352..5382
 /rpt_type=tandem
 complement(<5667..>7391)
 /locus_tag="PF14_0409"
 complement(<5667..>7391)
 /locus_tag="PF14_0409"
 complement(5667..7391)
 /locus_tag="PF14_0409"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AA037022.1"
 /db_xref="GI:23497480"

translation="MFNYFSAFLDNNINKNDNVLGKTLTNANNENERTKNLPLKD
 RLVANVQNLKNGINKNTDENSEKTLANNFTKPTIGIENTPTIINBTFNRKQY
 SIATVSNKQNEIEKKLKKRKKKHVILMERKQEKREKQKQMDQDQGEQYDDE
 EHEEEDNDNDYQDQNEOEDEHFEDDYEAENDNNNNNNNNNNNINRSFKNRS
 KQKDENDPNHKKQKGDDEYEDVDDSDNDQSQMKPKRKKCKKMEYERARLKEKHQ
 YHSGNVNPTFHTTSSNNHNSDSSQMKPKRKKCKKMEYERARLKEKHQ
 QNSLSILNTKNSFNKNDNNSTQSEIKVKRRGRPKASEVAAINNNKLNKQNDIRKY
 YKSGNDGFMKSNENKREKQNEEANTVFKIISKSTQENSFFMKQSPNFIIVLEK
 NYKNKIKGVTICPLINYQSKGAMAILIGDTTYGPKNSYGFMTFLVLDCHTNNFF
 IDTGIKNVIECEKHQILLISPGDMYFMQKQMEARLILVCKNNQNPFDVQKHIX
 DPMQNEK"

7729..7782
 /rpt_type=tandem
 7979..8020
 /rpt_type=tandem
 8386..8438
 /rpt_type=tandem
 9063..9109
 /rpt_type=tandem
 9264..9293
 /rpt_type=tandem
 <9317..>11238
 /locus_tag="PF14_0410"
 join(<9317..9689,9836..>11238)
 /locus_tag="PF14_0410"
 join(9317..9689,9836..11238)
 /locus_tag="PF14_0410"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AA037023.1"
 /db_xref="GI:23497481"

```
/translation="MRKLLSTALMDLKKIKNNPESKXNDNLSKLIIRNIHKFENDE
ILKIERKNEYKDTIKLCSYDFVFNHRSYFQDINKIRLYNISEIYDIHFNTS
VFNVIIRKLDIAIPYVNVFNRDYNINILKEVEIKIRIKNFNLDLTIIISFTI
LOELLKKI PVNNKSVENIYDTKDHSLCSALYDMYDKIILKIKQKKNH
LSVNSVLLNMLSRINFENYIEIKFTKYNKLNLDYDLEPHFILLNSPAKNH
INIMKYLKMNKNFENIINISYVNIINAVHYMAKFNATFLNLKOKVIELIILP
QRESNIMWSLAKUHIKDDFYIYAFQIKQIDVMMMSVAQLDMMRRKNSNGQ
DIGATIEKKLENDISCPUYDIKDTLNGONIDYKNNIBHKNHNEEKVEKEIIT
TQTTQKNDLNNIQLTYKPENNIISIXNTINTYISDLEKNIHLVLVKNYIKHI
QCHSLHVLTVQVFCQLGNTINSIYVKSLEILRKENDMTLNLIVAKYFLRIFIEK
QEAHFQKLPRSLKQFAKILNSDNN"
9716..9811
/rpt_type=tandem
repeat_region
11420..11447
/rpt_type=tandem
repeat_region
11803..11849
/rpt_type=tandem
repeat_region
11893..11918
/rpt_type=tandem
repeat_region
<12051..>12703
/rpt_type=tandem
repeat_region
locus_tag="PF14_0411"
mRNA
join(<12051..12111,12315..12394,12539..>12703)
CDS
join(12051..12111,12315..12394,12539..12703)
locus_tag="PF14_0411"
/codon_start=1
/product="small nuclear ribonuclear protein, putative"
/protein_id="AAN37024.1"
/db_xref="GI:23497482"
/translation="MATISGETFLPLALMDKICIGSKIWMKGDKEIVKLVGDFEY
VNVLEDEYETVYANNVKNYKIKKLLNGLNITIMVPGTGVNYYVDEKLEBNA"
12495..12537
/rpt_type=tandem
repeat_region
complement(<12999..>15647)
gene
locus_tag="PF14_0412"
mRNA
complement(join(<12999..15334,15524..>15647))
CDS
complement(join(12999..15334,15524..15647))
locus_tag="PF14_0412"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAN37025.1"
/db_xref="GI:23497483"
/translation="MNPDCDTRKFRDISEKIHGKYVPIIFKGGSKYIKNDYKNEADI
NNISYASFLNNVDIYDNKYGINKVCFSPQKIVAGTSGNFIYLYDLYENKLLSK
ICTRIDIOIADSDNKIYACENRLIEIFDIENKRCNMDIGKVDSVPSII
KNTLKNKCSNDSNKHNNILSKNLDNKYIPYHKYKINSKGSIVPIYNYED
LIKSNIMDLNLYLKNKNEIINKSFELIKESHEVITSCLAIPIPCNELVYSGGDG
VLKDWLRMLNFIKSKSSNSKSDIIFDNKPPFASICSHEDIVTSITFNKIIDHEG
SYLKQKQKDKKIFNSKSCIDNKNVYSSSLPCENNDSTSDYEDRESNVSPP
HSSICYSGLTFDLSKDFENILMTSGYDGIYRILYISNNIKSFYDEBSIITHCFSN
NNKIIISITNSKFKQKDFMVMNNKSNMKNATYANNKYTYCLNKLIRKENLENY
HDTINDDIFSSQYNDKVALCKIGNIPEDELEERIKIIVNELNRKINEQNKPEP
NKNDNNNNNNNNNNKNDKKNKYKVDNNKYNNNDNNITVEHDDYVESNYS
NSDSIYDENTIKSTYEVHVNKLEPTWSLFDVNLKYMDLVDPYDINISLKNYNYV
KSYFNNHNLNTINSKDEPLSYLMSEIISNEDDIPFSSCIAGDKICLIPSID
TYAHVYDLYTGFHVSINKNLYPNYVYDNHSSYSHSIDTNKKRISFLTSVDTYPKNQ
IIATNSNGYPPGSIYVAFAR"
13628..13761
/rpt_type=tandem
repeat_region
15450..15497
/rpt_type=tandem
repeat_region
15809..15855
/rpt_type=tandem
repeat_region
complement(<17033..>19621)
gene
locus_tag="PF14_0413"
mRNA
complement(<17033..>19621)
CDS
complement(17033..19621)
locus_tag="PF14_0413"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAN37026.1"
```

```
/db_xref="GI:23497484"
/translation="MSSVFNFYFKNFLVKEKNLFFKYNILERYSKITSVNLNNMN
IHKIIIOKINNSDFVSDIVETVGLHLDKERVISLDSYEAHCYCAKSPFCQIGISA
KKPDIVELYOTDENNKIKKKSVMHTRKVIYENGNNKICNIEKKQBNHISPYCIY
VPEKNKPSVSTSLSPLENEDFNEMIPNGVGLRPNNEBKKNIVKELNELNN
LLSKNTNNKNNIIFGEKIHDKIIQIEIENKIVUDEDEKALISIIKIKGWINNTND
Query Match 3.2%; Score 53.2; DB 3; Length 249943;
Best Local Similarity 43.4%; Pred. No. 0.32;
Matches 294; Conservative 0; Mismatches 383; Indels 1; Gaps 1;
QY 684 ACATTCACATTTTCGTTTGTAGCCTTTCTTTCTGGTGCTTACGGTCCCTCTTCCTGGT 743
Db 190305 ATATACCTCCTGTGCATATATATAAAAAATATGTTGAATACATAAAATATCTTTATCTACA 190246
QY 744 CGTCGTATGTACAAGTAGCATAGCTAGTGTGTTCCAAACCCGAAACAAAGTACCAACGAATC 803
Db 190245 CATATAATCTTCACATTTAATTAATAAAAAACAAAAACAAAAACAAAAACAAAAACAA 190186
QY 804 AAATAAGTTTGAATCGGTACATCTAGTTACCGTCCGACCTTACAACTTCAATCATTCGATTACT 863
Db 190185 AAAAAACAAATAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 190126
QY 864 TTGATCTCATTTCTAGTTTCGGTGTGTTATGTTTAAATATCCGATTTGTCAAGTACACAAGT 923
Db 190125 ATGTTTTTTTATATATATCTATTTTAAATATATATATATATATATATATATATATAGGT 190066
QY 924 ACATAAGTATCGGTATATGATGACCGGTTTAAATCAAGGACGAGTAGGAGGATT 983
Db 190065 ATTGAATTTATATATTTCTTTGAAGAATTAATAAAAAACAAAAACAAAAACAAAAACAA 190006
QY 984 TTGGAATCCTCGAAGAGGATTATTCATAGACACTAATTAGCTTTTGTGGCGCAGCC 1043
Db 190005 TATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 189946
QY 1044 TTGTGACCTACATTAATGGGGTCCAAACCCCAAGTATGGGCTTACAGCTTTTTCATATAAA 1103
Db 189945 GACAAAACATATATATTAATTAATCATATATATTTGTAAACCTGTGTATAATTAACGATAAAC 189886
QY 1104 TTAAGTAATCTTTTGTGCTA-ACCAATAAAATATTGAAATCTTTTCCAAACCAT 1162
Db 189885 AAAAAATATCAATATGTTTATATATATATATATATATATATATATATATATATATATATAT 189826
QY 1163 GAAAGTTTAAATTTGATCAGCGATGGAATTTTGTACAAAGCTAGGTATTTTCATTGGG 1222
Db 189825 ATTATTAATTTGTTTGAATTTATTTTAACTCTTAATAGTAACACGTCACAACTCTTT 189766
QY 1223 AGTGACTAGTAACAGTACTAACCAGATGATTTCTGATTTTGGATTTGAGC 1282
Db 189765 TTTATATATTTTATATATATATATATATATATATATATATATATATATATATATAT 189706
QY 1283 TTTTCTTAGGTTAAAAACCAAGTATTTTCTAAACATAAAAGAAAAACATTTTGTGAAA 1342
Db 189705 CTTTATTTTATATATATATATATATATATATATATATATATATATATATATATAT 189646
QY 1343 AGAGAAATAAGTTTACT 1360
Db 189645 TAAAAATTAATGTTTATT 189628
RESULT 12
AC117081
LOCUS
DEFINITION Dictyostelium discoideum chromosome 2 map 4903137-4960091 strain
AX4, *** SEQUENCING IN PROGRESS ***.
ACCESSION AC117081
VERSION AC117081.1 GI:20066271
KEYWORDS HTG; HTGS_PHASE2
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
REFERENCE Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.
AUTHORS Gloeckner,G., Eichinger,L., Szatranski,K., Pachebat,J., Dear,P.,
AC117081 56956 bp DNA linear HTG 24-MAY-2002
Dictyostelium discoideum chromosome 2 map 4903137-4960091 strain
```

Lehmann, R., Baumgart, C., Parra, G., April, J. F., Guigo, R., Kumpf, K.,
Tunggal, B., Cox, E., Quail, M. A., Platzer, M., Rosenthal, A. and
Noegel, A. A.
Sequence and Analysis of Chromosome 2 of Dictyostelium
The Dictyostelium Genome Sequencing Consortium
2 (bases 1 to 56956)
Baumgart, C.
Direct Submission
Submitted (06-APR-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
3 (bases 1 to 56956)
Baumgart, C.
Direct Submission
Submitted (24-MAY-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
CDS predictions from GeneID may contain errors. Further Information
is available from IMB Jena, Department of Genome Analysis
(http://genome.imb-jena.de/dictyostelium/)
and the University Cologne, Institute for Biochemistry I
(http://www.uni-koeln.de/dictyostelium/project.shtml)
Funding
Agency : Deutsche Forschungsgemeinschaft (DFG).
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1..56956
/organism="Dictyostelium discoideum"
/mol_type="genomic DNA"
/strain="AX4"
/db_xref="taxon:44689"
/chromosome="2"
/map="4303137..4960091"
join(33..523,639..2638)
/note="ORF_ID:dd_02927"
/pseudo
/codon_start=1
join(3306..3316,3871..4174)
/note="ORF_ID:dd_02929"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM43758.1"
/db_xref="GI:21166141"
/translation="MIVCILSTSFSSKTHSTAFENDLADSEIWEVNIISMLP
KFNVKGQVQLELIEGVNQSWNSKIGNVNIIYRTNLTFFPDHLSNRSDSIE
KN"
join(4284..4740,4840..5898,6096..6795,8057..8351,
8472..8483)
/note="ORF_ID:dd_02930"
/pseudo
/codon_start=1
join(8354..8906,9008..11010)
/note="ORF_ID:dd_02932"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM43759.1"
/db_xref="GI:21166142"
/translation="MSNGPSSSNRTTRKHKHIDISNITGTDSPCOLIIVTN
NSMGLPTCTCFPEINLERYFIGNFINLNTSSPSIIPNIVFNNTSIEGS
IVFYELLFGENLGSWFFFAVSDSVGTWMIEMPSLFLFCNGEFPFLNLSY
SKSNVYTFUGLGPILNLDWQSDTLVDIGSFTYNSIINLTINENCNGVN
TNFYIKTLSENVDQSLIEDIISYTNGLITQTLNPGVINLTINENCNGVN
YCLSDIGKCDPQHGGDLSPIELCNDNVCQNNOTGICSPISPYEWSGIDC
SILPHITSVSPDSGASISIGWNNHTYLOVFIGNKQCIPIYNISKEBIIICAP
PSTGLSVNQSINVTSDIYOYNDKQCPNCHCTSTNGICNTTGCYCGIGWS
GYDLSLVNPSGGDNGSGLPSTTDGNGTNVNTINQNTYQNTLILKLEIDFNGN
QINELLNKLWISNLSITVTYTIQIQNDKNTIITIDEINKRDITTFAGINFKUT
SGVMKTVSIENYYSNLTQLQMSVNEIQTNNENDCNNDQIEINTPNQDSL
NVTIKADKNGRFLIDRIESDGIPTMOTLVKSDSITVGMPLPHCKLCLDLP
DFSVLVSPFVDECGDNRKSWFLPVVIVPVIIVFLTLIIIFVITIKRFEVWFLPK
SKIKQLANLNK"

join(11974..12125,12329..12522,12630..13054)
/note="ORF_ID:dd_02935"
/codon_start=1
/product="Putative GTPase"
/protein_id="AAM43760.1"
/db_xref="GI:21166143"
/translation="MDPLSISWESKVVLLGSSDVGTALSLRYVDGIPPKELTSTIGA
SFLTTINIEGKIKYLIWSDAGQDRFSLATLYRGACVAILYVDITQKTFDVKG
WYELKANIQBEIIVLCNKIDLEQNRVSEKAYADEINAMVYETSKEGVE
GMFLIEGKGLILNKHNEYKFOQQQKLKFOQQQNSLYQYHIQQQKHQQQQQQQ
QIFKQQFYNNGH5LQGSINHNQNSTYNSDSDCCG"
complement(join(13444..14192,14284..14623,14713..15015))
/note="ORF_ID:dd_02938"
/codon_start=1
/product="AMMONIUM TRANSPORTER AMTA"
/protein_id="AAM43761.1"
/db_xref="GI:21166144"
/translation="MVAGEIIGVAAEITNGSSSVQVYLDCAQVAPDGNTTWVL
LSTILVGMPLAFFEAGLLRSNXTLSIITQMSGIVLTVWQAFYSLTGPDOK
GIIGNLDAFLINYSYDDCSNAPNIIPAAVAFMMFMANITPLMTGAPAEYKFKFA
FIALTVAWEIIVFYVAHWINGWGLHYFGLVDPAAGIVHTSAGVSALVIALYVGR
RDFEYGGEPFNNLPLATIGALLMWWFGFNAGSALAAGNATSAVASTQIGGSF
SAIVMILSAAKGPNTSVINGVAGLITPASGYINSQYSGIGCIGLASIYVS
VLLKHLHIDALDVSVHGLTGIIGSLAIGCAELSVNPNANGAFYGNPLGTOL
LCVSVAVMAAAWTVLLKIIDATIGVKIDSEBELGLDVEHGEFAYHNIISLQGEN
HYSSVINSHDFPK"
15832..17823
/note="ORF_ID:dd_03054"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM43762.1"
/db_xref="GI:21166145"
/translation="WKDILSPKYLHIDILKPIIDYLLTNNKINYNDDNNNNKVIK
NYLYINEILKLSLVSHWYQIISNLISNLNLEEWPDLYSDVPEENYKGI
NFIAYDFNKIYDVITQSNFKQVQKSIINKQSVIEDFFRNQVLCIISDNTVEKLL
QLSKLETGNLYKRIKSNPKYSGKGNLPGEGIAKGIEKXGLDIDSSNNNNNN
NN
YDIYDQPIILYNTIDYEGSDDYNEKTRILYKLSKNIIPRSNTNENVHADFSLM
FNNKINNOVESIRIDSFVPIDLQNVVELTNLHTLSIPLHFHOMILLQIGLHCR
GGDLEHKNQVMDKMWIRSLLSAKSNKPKSNKCYTECNKSINSYNNENYSA
SRHVKLPNGIKQLSSSTFKFNSMYDIIISDIFSELKNNKTKLKYVKEKPHY
GIRLFTTSERVAEYQFNINLYIENFSGSGSGLSNKIRKFRFSLAFPPYKRLI
FESLPKHYNYFOLYSITVIYDTELLNEIINLSKISKIVQOQQOQQOQSIKEFKI
VIERKFSNKHQFQILANDRINGDGLFVITVSNKKELC"
complement(join(17912..18092,18308..18333))
/note="ORF_ID:dd_03055"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM43763.1"
/db_xref="GI:21166146"
/translation="WTIIFLICLDASTQSTTNNNSINNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNSKVDFNIE"
18436..20349
/note="ORF_ID:dd_02950"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAM43764.1"
/db_xref="GI:21166147"
/translation="WKDILSPKYLHIDILKPIIDYLLTNNKINYNDDNNNNKVIK
NYLYINEILKLSLVSHWYQIISNLISNLNLEEWPDLYSDVPEENYKGI
NFIAYDFNKIYDVITQSNFKQVQKSIINKQSVIEDFFRNQVLCIISDNTVEK
LLQSKLETGNLYKRIKSNPKYSGKGNLPGEGIAKGIEKXGLDIDSSNNNNNN
NN
YNEKTRILEYKSNKTRFTSVNHADPSFLMNNKINNQVSEIISDFDVEPDL
QNVVELTNLHTLSIPLHFHOMILLQIGTIELHCRGGDLEHKNQVMDKMWIR
SKSIKNPKIYNKVCIGCINQSIKSYNNENYSAASROVKLFSQIGQLSSSTFKP
KYNNHDIISDIFSELKNNKTKLKYVKEKFLYVCLRNKEVVAHQFINIYLIEN
VFSGASGSLSNKIRKFRFSLAFPPYKRLIIPESLPKHYNYFOLYSITVIY
ELLNEIINLSKISKIVQOQQOQQOQSIKEFKIIVINKINNKHQFQILANDRINGD
GLFVITVSNKKELC"
join(22336..23207,23308..23806)
/note="ORF_ID:dd_02928"
/codon_start=1

/product="rac-family serine/threonine kinase homolog"
/protein_id="AA043765.1"
/db_xref="GI:21166148"
/translation="MTTISNLPOLSEISPTKLLITKQPIKRESTSPKSWMDTP
KNTIISPTCESESSITTPSESLSGHESSISDSKTIITAILLAPNSVQ
DFTPLCNGAETIIVKISINLSNFIIKELVKGFGKVFQVHVDTQKVALKVIK
KNHIIAASVNTFAELIKKISHPFIIVLHYAFQNEKXLYLMDVNGQGLFHLQ
KEAIFSEDQVRFYAEILIALEHLSDNIHVRLDLKPNILLDSQGHILITDFGLAKLE
VKTNTTSPFAGTLYEMAPEMICHATGKADWMSIGILMYDMWIKGPKPFHKNRALM
QEKIIEKAFKPFVSSARSLLNGLLTKDPTKRLGANGAIEIKRHPFKSIQWKEIE
NKEITPPFVSTGIDDISNFDHASLKAQRDSFSTSLSSQQAYPDGFSFVTPV
LLESQ"
complement(join(24089..24410,24517..24779))
/note="ORF ID:dd_02931"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA043766.1"
/db_xref="GI:21166149"
/translation="MSDPGEENVEITEFVEGDINENDLIDGNEYVDNGISPEITT
TFDNNNNNNNNNNNSYSGDGLSSVDGDMKPEAPAPVRYLKEHEKMEKK
KXSEKQKQKIAKQSLDNFYSEBAKKTALKNNRDNHKSLETDSTSGNTHTWES
VWSMIDLAQPNPANKDTERMREILRLKNQPIV"
25279..26412
/note="ORF ID:dd_02934"
/codon_start=1
/product="Mitochondrial genome maintenance protein MGM101
precursor."
/protein_id="AA043767.1"
/db_xref="GI:21166150"
/translation="MKRFISTSLRLVSLNKNNNNNKFNKSIIVGFNREFSSNEG
LNKVFNGNNNSSTTPTKVLNDSQYQEDKYQISKEPPSKIEIVDTLLADLNDDI
EIKPDGLIYLPKYLRIINCAEPGQWALKPFPVVGKTLIRPVALYLCIRVVAE
SIGEQVFNPSFISFATASAKNALVRCCKDLGSSLMDFPIRQWKSIAIYERW
CENSKTKERLFLVNRNSQULPFWKESDFNFQVPSNYSISSNNNNNNNN
NNNSDSSSINQOQETISYHDSQSSPSKITEQDDSEIDIDDVWPPQLKYAGK
TWQLVADPKGLSVLWASSSFDNAPKIKSQATAILEFINKSQ"
complement(26482..27039)
/note="ORF ID:dd_03240"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA043768.1"
/db_xref="GI:21166151"
/translation="MFIYEPFHHVNTAIPDFKILTFVQVFNITIIIPCKEFLSKDTVT

Query Match 3.1%; Score 52.6; DB 2; Length 56956;
Best Local Similarity 46.6%; Pred. No. 0.54; Indels 0; Gaps 0;
Matches 169; Conservative 0; Mismatches 194; Indels 0; Gaps 0;
QY 1091 TTTTTCATATAAATAAGTAAATCTTTTTCCTAACCAATAAAATTAATGAAAATC 1150
DB 52626 TTTTAACTATAATTTATTTATTTTGTGTTTATTTTAAATTTATTTTAAAT 52685
QY 1151 TTTCCAACTAGAAAAGTTAAATTTGATCAGCGATGGAATTTTGTCAAAGCTAGGT 1210
DB 52686 ATTATTTATTTTAAATTTATAAAATTTAGAAAATATAAAAAAATAAAAAACACAC 52745
QY 1211 ATTTTCATTGGGAGTGACTAGTAAGTAACTAACCAAGTACAGTTTCTGATTTT 1270
DB 52746 TTTTAAATTAGATGGGATTTATTTATTTAGAGTTTATTTATATATAAAAGGGGTGT 52805
QY 1271 GGATTTTGAAGCTTTCTTAGTTTAAAAACAAGATATATTACTAAACAATAAAGAAAAA 1330
DB 52806 CCAAGTAAAGAGTTTATTAATGTAATGAAAAATAAAAAAATAAAAAATTTTAAATAAAA 52865
QY 1331 CATTTTGTGAAGAGATAATAAGTTTACTTGACCCCATTTGTACAGATGGTCCCATATA 1390
DB 52866 AATGTAAACAAAAAATAAAAAAATAAAAAATAGATGTGGGGATAGAAATAAAAAA 52925
QY 1391 ATACTGTAGAGATAGACGAATGGAAGTGAATTTTGTTCAGTGGTACATCGGAATGGT 1450
DB 52926 AAAAAATAAAAAAATAATTTTATTTTATTTTGTGAAATATCAATTTTAAATTTT 52985
QY 1451 TCT 1453

DB 52986 TTT 52988
RESULT 13
AX461144/c
LOCUS AX461144.1 2000 bp DNA linear PAT 08-JUL-2002
DEFINITION Sequence 73 from Patent WO0198480.
ACCESSION AX461144
VERSION AX461144.1 GI:212726352
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1
AUTHORS Budworth,P., Brown,D., Chang,H.S., Zhu,T., Han,B., Wang,X. and
Cooper,B.
PROMOTERS Promoters for regulation of plant gene expression
JOURNAL Patent: WO 0198480-A 73 27-DEC-2001;
SYNGENTA Participations AG (CH)
FEATURES
1..2000
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
BASE COUNT 714 a 371 c 305 g 610 t
ORIGIN
Query Match 3.1%; Score 52.4; DB 6; Length 2000;
Best Local Similarity 47.1%; Pred. No. 0.98;
Matches 161; Conservative 0; Mismatches 181; Indels 0; Gaps 0;
QY 1075 AGTAGGGCTTACAGCTTTTCCATAAAATTAAGTAAATCTTTTTCGCTAACCAATA 1134
DB 1019 ATTAGGTCGTATCTCTTTAACTATAAAAAAATAAACTATATGATGATGAAATG 960
QY 1135 AAAATATTGAATCTTCCACCATAGAAAAGTTAAATTTGATCAGGATGGAATTT 1194
DB 959 AGAAAAATGAAGTACGAAACAGGTCTTAAACTGTTCCTATTTTATTTGATGATATCAT 900
QY 1195 TTGTACAAAGCTAGTATTTCATTTCGGAGTGTACTAGTAACTAGTAAAGTAAACAGA 1254
DB 899 TTCAACTCTGAGATTTATTAGCATTAACCTACACGAAAAAATAAACTTGCATTAC 840
QY 1255 ATGAGTTCTGATTTTGGATTTTGAAGCTTTTCTAGGTTTAAAAAACAGTATATTACTA 1314
DB 839 AAGTGTTTTTTTTTTTCTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT 780
QY 1315 AACATAAAAGAAAAACATTTTGTGAAAAAGAAAAATAAGTTTACTGGACCCCATTTGTAC 1374
DB 779 GCAAGCGAAGCTACCTCTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 720
QY 1375 AGATGGTCCCATATAATATCTATGATAGAGATAGAGCAATGGA 1416
DB 719 GCCCTGACGGTTAATCGCATCGGTAGAGAAAGATCGGTGGA 678
RESULT 14
AC004401/c
LOCUS AC004401.3 GI:20197125
DEFINITION Arabidopsis thaliana chromosome 2 clone F21P24 map C1C06C07,
complete sequence.
ACCESSION AC004401
VERSION AC004401.3
KEYWORDS HTG.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 94038)
REFERENCE 1 (bases 1 to 94038)

AUTHORS Rounsley, S.D., Lin, X., Ketchum, K.A., Crosby, M.L., Brandon, R.C., Sykes, S.M., Kaul, J.S., Mason, J.M., Kerlavage, A.R., Adams, M.D., Somerville, C.R. and Venter, J.C.
Unpublished
2 (bases 1 to 94038)
Lin, X.
Direct Submission
Submitted (09-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
3 (bases 1 to 94038)
Town, C.D. and Kaul, S.
Direct Submission
Submitted (27-FEB-2002) The Institute for Genomic Research, 9712 Medical Center Dr. Rockville, MD 20850, USA, cdtown@igir.org
On Apr 18, 2002 this sequence version replaced GI:16598413.
FEATURES
Location/Qualifiers
1..94038
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="2"
/map="C106C07"
/clone="F21P24"
1..22544
/note="overlap with BAC clone T20K9 (AC004786:64143..86686)."
1..1327
/gene="At2g22950"
/note="synonym: F21P24.1"
join(<1..92,168..326,400..571,679..977,1088..>1327)
/gene="At2g22950"
join(1..92,168..326,400..571,679..977,1088..1327)
/gene="At2g22950"
/note="protein sequence is in conflict with the conceptual translation"
/codon_start=1
/product="hypothetical protein"
/protein_id="MAF18608.2"
/db_xref="GI:20197127"
translation="TLTDDGTAIGPVPERKQREMLELIPKIQVMASSPMKHTLV
KQRTTFDEVNAGVGDNDAPALHEADIGLNGAGTGEVAKETADVILDDNFSTLV
TVAKGRSVYINIQVQFQTVNVVALVNFSACTGSAPLTAVQLLWVNMIMDTL
GALALATPNNLEMPVGRGNFTNAMRNILGQAVYQFIILWLOAKGKSMFG
LVGSDTLVNTLFCFVFCQVNEVSSREMEIDVFKGILDNVFVWIGATVFFQ
IILIEPLGTFTSTPLTQVWFSSIFVGLGMPAAGLKIPV"
complement(594..665)
/rpt_family="(TAAAAA)n"
repeat_region
1891..1968
/rpt_family="AT_rich"
2455..4493
/gene="At2g22960"
/note="synonym: F21P24.2"
join(<2455..2582,2843..2857,3096..3156,3867..3985.
4057..4170,4277..4326,4417..>4493)
/gene="At2g22960"
join(2455..2582,2843..2857,3096..3156,3867..3985,
4057..4170,4277..4326,4417..4493)
/gene="At2g22960"
/codon_start=1
/product="putative serine carboxypeptidase I"
/protein_id="AAV14928.1"
/db_xref="GI:20197126"
translation="NRTSPKLLLLLVLRHHAESGIVKFLPGFPGPLPFELETGY
IGIGLAKYWANDERVALQIRKSGIKWRCNSNIHDDDISSIPYMNNSINGY
RSLIYSGDHDMVEPPLATEAMIRSLNYPILDDRPWLNQIAGYTWYANKQTYATI
KASGHTASYKPAEGTFMQVISOQL"
complement(3355..3405)
repeat_region
4553..4583
/rpt_family="(TA)n"
repeat_region
complement(4633..4653)
/rpt_family="AT_rich"
repeat_region
5266..5297
/rpt_family="AT_rich"
6421..10087
/gene="At2g22970"
/note="synonym: F21P24.3; supported by cDNA:
gi_14517521.gb.AY039596.1"
join(6421..6550,6635..6780,6881..6957,7038..7160,
7365..7667,7736..7773,7979..8077,8179..8265,8358..8426,
8612..8683,8764..8885,9346..9459,9572..9617,9705..10087)
/gene="At2g22970"
join(6436..6560,6635..6780,6881..6957,7038..7160,
7365..7667,7736..7773,7979..8077,8179..8265,8358..8426,
8612..8683,8764..8885,9346..9459,9572..9617,9705..9785)
/gene="At2g22970"
/codon_start=1
/product="putative serine carboxypeptidase I"
/protein_id="AAC17814.1"
/db_xref="GI:3169171"
translation="NELTLKLLVLLFNLHHVSGSIVKFLPGFPGPLPFELETGYI
GIGEEVQLFYFYFKSRNPKEDPLLWLSGGPCSSITGLFONGPLALKSEVYNG
SVPSLVSTYTSWTKTANIIFLDQVAGFSYSRAPLIDTPTDTGEVRAIHEFLQKWL
KIPSSNNHFYAGDSYGMIVFVAGISGNYICCNPPINLGVIGUNPITHEDDP
NTRIIFSGMALISDELYSIREACKNYFNVDPRNTKCLKVEEFKCTDKIAEFHI
LSPDCDTASPCYLYFYFYLISFANDSVRDALHYNKRISIGKVERCNLYSPKINDKI
SSVPYMNNSVSGYSLIYSGDHLVFPFLATQAWIKSLNYSIIDWRPMIRDIQTG
VYRTYSNMWTRATVAGSGHTAENKQESFINFRWINGQL"
7277..7392
/rpt_family="AT_rich"
9116..9138
/rpt_family="AT_rich"
10590..14726
/gene="At2g22980"
/note="synonym: F21P24.4; supported by cDNA:
gi_15294269.gb.AF410326.1 AF410326"
join(<10590..10717,10800..10945,11012..11088,11168..11290.
11602..11704,11806..11828,11912..12013,12378..12464,
12552..12620,12727..12798,13609..13730,14190..14303,
14407..14452,14559..14726)
/gene="At2g22980"
join(10590..10717,10800..10945,11012..11088,11168..11290,
11602..11704,11806..11828,11912..12013,12378..12464,
12552..12620,12727..12798,13609..13730,14190..14303,
14407..14452,14559..14639)
/gene="At2g22980"
/codon_start=1
/product="putative serine carboxypeptidase I"
/protein_id="AAC17815.1"
/db_xref="GI:3169172"
translation="NSLTLEFLLLVILSHHAHSGSIVKFLPGFPGPLPFELETGY
IGIGEEVQLFYFYFKSRNPKEDPLLWLSGGPCSSITGLFENGVALKFEVYN
GSVPISLVSTYTSWTKTANIIFLDQVAGFSYSRAPLIDTPTDTGEVRAIHEFLQKWL
SRHQQFFNPVYVGGDSYGMIVFVAGISGNYICCNPPINLGVIGUNPITHEDDP
IPVAGMALISDELYKSVERICKNYKVSNTKCYLIKDQYCKHLKAKYHLLP
DQDITSPCFLYRYTLITFWANNKSVREALQVNGSGIKWYCKYKALSYNDIKSSV
ATHMNSIDGKSLYINGHDHDMVFFLATQAWIKSLNYSIIDWRPMIRDIQTG
SYSNMWTATKSGSHTAEPKETSIMFKWISAQPL"
complement(11836..11877)
/rpt_family="AT_rich"
17236..17266
/rpt_family="(TAA)n"
17877..21740
/gene="At2g22990"
/note="synonym: F21P24.5; supported by cDNA:
gi_14334757.gb.AY035052.1"
join(17877..18072,18144..18289,18426..18502,18632..18751,
19006..19108,19205..19242,19338..19439,19871..19957,
20049..20126,20326..20397,20476..20594,21039..21152,
21262..21311,21410..21740)
/gene="At2g22990"
repeat_region
repeat_region
repeat_region
repeat_region
gene
mRNA

| | | |
|-------------|-------|---|
| CDS | | join(17954...18072,18144...18289,18426...18502,18632...18751, 19006...19108,19205...19342,19338...19439,19871...19957, 20049...20126,20326...20397,20476...20594,21039...21152, 21262...21311,21410...21486) /gene="At2g22990" /codon_start=1 /product="putative serine carboxypeptidase I" /protein_id="AACI7816.2" /db_xref="GI:20191728" /translation="MSLKIKFLLLVVHHVDASIVKLFPGFGPIPLPELETGYGIG GEDENVQFYFTYKIWIWNGSGCGSLGIIIFENGVPGLKFVNCSA PSLFTGTSTYXMANIIIDQVGSGFSYKPIDKTGDISEVRTHFEFLQWLRSRP RIPYAVGLGISDELYEMKRICNGNYNVDSPNTQLKTEEVHKCTAKINIHILT PCDVTNVTSPPCYYPFHLEICWANDSVREALHIEKGSKGWARNRIPTYNHDIV SYIPHMNSISGISYLSGDIAVFATQWIRSLNYSPIHNWRPWNINNIAG YTRAYNMOTFAIKASGHATABYRNFTFMFORWISQQPL" complement(18316...18358) /rpt_family="(TA)n" repeat_region 18578...18613 /rpt_family="(A)n" repeat_region 19798...19848 /rpt_family="(AT_rich" complement(20808...20858) /rpt_family="(AT_rich" repeat_region 21369...21390 /rpt_family="(AT_rich" repeat_region 22307...22327 /rpt_family="(A)n" gene 23773...27421 /gene="At2g23000" /note="synonym: F21P24.6; supported by full length cDNA: Ceres:33165" mRNA join(23773...23969,24056...24201,24279...24355,24441...24560, 24831...24933,25023...25060,25156...25257,25439...25525, 25617...25700,25899...25970,26085...26203,26875...26988, 27084...27129,27222...27421) /gene="At2g23000" CDS join(23845...23969,24056...24201,24279...24355,24441...24560, 24831...24933,25023...25060,25156...25257,25439...25525, 25617...25700,25899...25970,26085...26203,26875...26988, 27084...27129,27222...27302) |
| Query Match | | 3.1%; Score 52.4; DB 8; Length 94038; Best Local Similarity 47.1%; Pred.No.0.55; Matches 161; Conservative 0; Mismatches 181; Indels 0; Gaps 0; |
| QY | 1075 | AGTATGGGTTACAGCTTTTCCTAAATAATTAAAGTAATCTTTTTTGCCCTAACCAATA 1134 |
| DB | 72682 | ATTAGTTGTAATCTTTTAACTATAAAAAAAAATAACTATATGAATGATGAATG 72623 |
| QY | 1135 | AAATTTATGAAAATCTTTCCACCATAGAAAGTAAATTTGCATGCGCATGAAATTT 1194 |
| DB | 72622 | AGAAAAATTGAAGTACGAAACAGGTCTCTAAACTGTTCTCTATTATTTGGTAGATATCAT 72563 |
| QY | 1195 | TGTACAAGCTAGGTATTCTTTGGGAGTCTACTAGTACTAGTAACTAGTAACTAACAGA 1254 |
| DB | 72562 | TTCAAACCTTGAGAAATTTTACATTAACCTCTACACGAAAAAAAATAAACTTCATAC 72503 |
| QY | 1255 | ATGAGTTCTCTGATTTTGGATTTTGAAGCTTTTCTTAGGTTAAAAAACAAGTATATTACTA 1314 |
| DB | 72502 | AAGTGTTTTTTTTTTTTTTTGAACACGTTAAAGAACCGAACTTGATCATCTATTATTA 72443 |
| QY | 1315 | ACACATAAAGAAAAACATTTTGTGAAGAGAAATAAAGTTTACTGCCATTTGTAC 1374 |
| DB | 72442 | GACAACGGAAGTACCCCTTTGTTGTGTCAGTCTCTATATAAAGTTTTTTCACGGAAC 72383 |
| QY | 1375 | AGATGTGCCATAATAATACTCATAGAGATAGCAATGGA 1416 |
| DB | 72382 | GCCTGACCGTTAATCGCATCGGTAGAGAAAGATCCGTCCA 72341 |
| RESULT 15 | | |

RESULT 15

| Query Match | Best Local Similarity | Score | DB | Length | Matches | Mismatches | Indels | Gaps |
|-------------|-----------------------|-------|------|--------|---------|------------|--------|------|
| Qy | 1034 | 76.1 | DB 3 | 34750 | 1093 | 0 | 0 | 0 |
| Db | 6842 | 76.1 | DB 3 | 34750 | 1093 | 0 | 0 | 0 |
| Qy | 1094 | 76.1 | DB 3 | 34750 | 1153 | 0 | 0 | 0 |
| Db | 6902 | 76.1 | DB 3 | 34750 | 1153 | 0 | 0 | 0 |
| Qy | 1154 | 76.1 | DB 3 | 34750 | 1213 | 0 | 0 | 0 |
| Db | 6962 | 76.1 | DB 3 | 34750 | 1213 | 0 | 0 | 0 |
| Qy | 1214 | 76.1 | DB 3 | 34750 | 1273 | 0 | 0 | 0 |
| Db | 7021 | 76.1 | DB 3 | 34750 | 1273 | 0 | 0 | 0 |

| | | | |
|----|------|--|------|
| Db | 7022 | TATATTAAATAAAATATATTATTATATAAATATATTATTATTTTAAATTTTGTGTTTTCT | 7081 |
| Qy | 1274 | TTTTGAGCTTTTCTTAGGTTAAATAAACAGTATATTACTATAACATATAAAGAAACAT | 1333 |
| Db | 7082 | TTTATTTCTTTAAATTAATATATTAAATAAATTTATTTACTATAATAAATTTATATCAT | 7141 |
| Qy | 1334 | TTTGTGAAAGAGAAATAAAGTTTACTGGACCCCATTTGACAGATGGTCCCATATAATA | 1393 |
| Db | 7142 | AAAATTAAATAAAGTAATACTATATAATAAATAAATTTATCTAAATCAACAAAAAACCT | 7201 |
| Qy | 1394 | CTGATAGAGATAGA | 1408 |
| Db | 7202 | TGGAACCAAAAAAGGA | 7216 |

Search completed: February 2, 2004, 00:53:01
Job time : 6372.59 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

M nucleic - nucleic search, using sw model

un on: February 1, 2004, 21:36:05 ; Search time 491.261 Seconds
(without alignments)
9214.976 Million cell updates/sec

title: US-09-938-842A-3729

perfect score: 1677
sequence: 1 ggttaagcgttttctctatg.....tttctctcagctatatattta 1677

scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

database : N_Geneseq_19Jun03.*

```
1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 1677 | 100.0 | 1677 | 24 | Arabidopsis thalia |
| 2 | 242.4 | 14.5 | 1483 | 21 | Arabidopsis thalia |
| 3 | 50.6 | 3.0 | 8056 | 22 | Haematopoietic cel |
| C 4 | 49.6 | 3.0 | 422 | 25 | Human breast cance |
| C 5 | 49 | 2.9 | 369 | 22 | Human cervical can |
| C 6 | 48.6 | 2.9 | 20933 | 24 | Human angiogenesis |
| C 7 | 47.8 | 2.9 | 2958 | 24 | Arabidopsis thalia |
| C 8 | 46.6 | 2.8 | 516 | 25 | Bovine EST associa |

| | | | | | | |
|------|------|-----|--------|----|----------|--------------------|
| 9 | 46.4 | 2.8 | 7047 | 24 | ABK28385 | DNA transcription |
| C 10 | 46.4 | 2.8 | 17848 | 22 | AAS45323 | Chemically pretrea |
| C 11 | 46.4 | 2.8 | 17848 | 24 | ABK39976 | Human chemically p |
| C 12 | 46.4 | 2.8 | 17848 | 24 | ABK28164 | DNA transcrip |
| C 13 | 46 | 2.7 | 8056 | 25 | ABZ10246 | Haematopoietic cel |
| C 14 | 45.6 | 2.7 | 37515 | 24 | ABQ66998 | Human angiogenesis |
| C 15 | 45.4 | 2.7 | 6352 | 24 | ABL70563 | Chemically treated |
| C 16 | 45.4 | 2.7 | 6352 | 24 | AAS61235 | Human gene regulat |
| C 17 | 45.4 | 2.7 | 6352 | 24 | ABK31340 | Signal transductio |
| C 18 | 44.8 | 2.7 | 1271 | 24 | ABZ15024 | Arabidopsis thalia |
| C 19 | 44.6 | 2.7 | 2000 | 24 | ABZ16233 | Arabidopsis thalia |
| C 20 | 44.2 | 2.6 | 7749 | 24 | ABL70435 | Chemically treated |
| C 21 | 44.2 | 2.6 | 8711 | 22 | AAS45700 | Tumour suppressor |
| C 22 | 44.2 | 2.6 | 19734 | 24 | ABL33933 | Human immune syste |
| C 23 | 44.2 | 2.6 | 113515 | 24 | ABL34175 | Human immune syste |
| C 24 | 44 | 2.6 | 468 | 23 | ABV07439 | Human prostate exp |
| C 25 | 44 | 2.6 | 6636 | 24 | ABN80114 | Human chemically m |
| C 26 | 44 | 2.6 | 7498 | 24 | ABL32257 | Human immune syste |
| C 27 | 43.6 | 2.6 | 10710 | 24 | ABN79996 | Human chemically m |
| C 28 | 43.6 | 2.6 | 11473 | 24 | ABK40030 | Human chemically p |
| C 29 | 43.6 | 2.6 | 11473 | 24 | ABL33355 | Human immune syste |
| C 30 | 43.6 | 2.6 | 14752 | 20 | AAX20256 | Borrelia burgdorfe |
| C 31 | 43.4 | 2.6 | 6533 | 24 | ABK28383 | DNA transcription |
| C 32 | 43.4 | 2.6 | 7008 | 22 | AAS46532 | Tumour suppressor |
| C 33 | 43.4 | 2.6 | 15161 | 24 | ABL70458 | Chemically treated |
| C 34 | 43.4 | 2.6 | 15161 | 24 | AAS61423 | Human gene regulat |
| C 35 | 43.4 | 2.6 | 640681 | 24 | ABA92787 | Buchnera sp. genom |
| C 36 | 43.2 | 2.6 | 6175 | 24 | ABL33307 | Human immune syste |
| C 37 | 43 | 2.6 | 6244 | 24 | ABL33485 | Human immune syste |
| C 38 | 43 | 2.6 | 7244 | 22 | AAS46726 | Tumour suppressor |
| C 39 | 43 | 2.6 | 40324 | 24 | ABQ67150 | Human angiogenesis |
| C 40 | 42.8 | 2.6 | 5682 | 24 | ABL32572 | Human immune syste |
| C 41 | 42.8 | 2.6 | 5682 | 24 | ABL34500 | Human metastasis a |
| C 42 | 42.8 | 2.6 | 7809 | 25 | ABZ03977 | Haematopoietic cel |
| C 43 | 42.8 | 2.6 | 7809 | 25 | ABZ10123 | Haematopoietic cel |
| C 44 | 42.8 | 2.6 | 8056 | 25 | ABZ10100 | Haematopoietic cel |
| C 45 | 42.8 | 2.6 | 8056 | 25 | ABZ10100 | Haematopoietic cel |

ALIGNMENTS

RESULT 1
ABZ15924
ID ABZ15924 standard; DNA; 1677 BP.
XX AC ABZ15924;
XX DT 21-JAN-2003 (first entry)
XX DE Arabidopsis thaliana stress regulated gene SEQ ID NO 3729.
XX KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX OS Arabidopsis thaliana.
XX PN WO200216655-A2.
XX PD 28-FEB-2002.
XX PF 24-AUG-2001; 2001WO-US26685.
XX PR 24-AUG-2000; 2000US-227866P.
XX PR 26-JAN-2001; 2001US-264647P.
XX PR 22-JUN-2001; 2001US-300111P.
XX (Scri) SCRIPPS RES INST.
XX (SYGN) SYNGENTA PARTICIPATIONS AG.
XX Harper JF, Kreps J, Wang X, Zhu T;
XX WPI; 2002-304127/34.

Identifying a stress condition to which a plant cell has been exposed and producing plants with increased tolerance to these abiotic stresses

Claim 144; SEQ ID NO 3729; 577pp + Sequence Listing; English.

The invention relates to identifying a stress condition to which a plant cell has been exposed, comprising:

(a) contacting nucleic acid representative of expressed polynucleotides in the plant cell with an array or probes representative of the plant cell genome; and

(b) detecting a profile of expressed polynucleotides in the plant cell characteristic of a stress response. The method is useful in the production of transgenic plants, cells and seeds and in producing plants with increased tolerance to abiotic stress. The present sequence is that of an Arabidopsis thaliana stress regulated gene (AB212196-AB217574) used in methods of the invention.

Note: The sequence data for this patent is not represented in the printed specification but is based on sequence information supplied to Derwent by the European Patent Office.

Sequence 1677 BP; 508 A; 292 C; 335 G; 542 T; 0 other;

| | | | | |
|-----------------------|---|---------------|-----------|--------------|
| Query Match | 100.0%; | Score 1677; | DB 24; | Length 1677; |
| Best Local Similarity | 100.0%; | Pred. No. 0; | | |
| Matches 1677; | Conservative 0; | Mismatches 0; | Indels 0; | Gaps 0; |
| 1 | GGTTAAGCGTTTACTTATGCTTATATGCAACGGAAGAAATATGCCATTGTTGGAATGC | 60 | | |
| 1 | GGTTAAGCGTTTACTTATGCTTATATGCAACGGAAGAAATATGCCATTGTTGGAATGC | 60 | | |
| 61 | TTTTTTCAGATCATCAAGGCTCTACAGATTTCTTAGGGAATGTTTCAGGCTTTTGTGA | 120 | | |
| 61 | TTTTTTCAGATCATCAAGGCTCTACAGATTTCTTAGGGAATGTTTCAGGCTTTTGTGA | 120 | | |
| 121 | GAATTTGTTTATTTGCAACAGCTAGAGAAATACCAATAGACAGATGTATCTGAAGAGA | 180 | | |
| 121 | GAATTTGTTTATTTGCAACAGCTAGAGAAATACCAATAGACAGATGTATCTGAAGAGA | 180 | | |
| 181 | TAACTCTCTATCTCTAAGAAATGACCGATACAGTAATCAACAGCATCATTAAGAT | 240 | | |
| 181 | TAACTCTCTATCTCTAAGAAATGACCGATACAGTAATCAACAGCATCATTAAGAT | 240 | | |
| 241 | TAAATGGTTTGTAAAGAAATCTACACTTATTTATGTGAAATTTGTTGGTTAGTGAAGAAGT | 300 | | |
| 241 | TAAATGGTTTGTAAAGAAATCTACACTTATTTATGTGAAATTTGTTGGTTAGTGAAGAAGT | 300 | | |
| 301 | AAAAACATCGGAATCCAAACCTCAATTTACCAATCAGCCCAAAATTTATGATGCTGGCG | 360 | | |
| 301 | AAAAACATCGGAATCCAAACCTCAATTTACCAATCAGCCCAAAATTTATGATGCTGGCG | 360 | | |
| 361 | TAAATGATGGTATGCTGATGTAGTGAAGAAAGTTGGTGGCTGCGAAATTTACACATTATC | 420 | | |
| 361 | TAAATGATGGTATGCTGATGTAGTGAAGAAAGTTGGTGGCTGCGAAATTTACACATTATC | 420 | | |
| 421 | CCTCTGGTGGACCGGAATCTGTAATTCGGAAGAGTGGAAACCCACCTTGTGTTTAACTTTT | 480 | | |
| 421 | CCTCTGGTGGACCGGAATCTGTAATTCGGAAGAGTGGAAACCCACCTTGTGTTTAACTTTT | 480 | | |
| 481 | AAGCTTAAAGGTTTACTACCGGTTTGACCGGTTTATTAATTTGTTTAACTTTTAACTTTT | 540 | | |
| 481 | AAGCTTAAAGGTTTACTACCGGTTTGACCGGTTTATTAATTTGTTTAACTTTTAACTTTT | 540 | | |
| 541 | GGATCGGTTTGTGTTAATCTCAAGGCCACGTTATCGCCAAATATTTGATTTTTCAGTG | 600 | | |
| 541 | GGATCGGTTTGTGTTAATCTCAAGGCCACGTTATCGCCAAATATTTGATTTTTCAGTG | 600 | | |
| 601 | GGTAGGGAATGGTGGGTGCAATAGTTGGGCTAGCCCTCAACAAATGTGGAACTGAAG | 660 | | |
| 601 | GGTAGGGAATGGTGGGTGCAATAGTTGGGCTAGCCCTCAACAAATGTGGAACTGAAG | 660 | | |
| 661 | AGAGTAGGCTCCAGCTCAGGCGCCCAATTCATTTTCGTTTGTAGCCTTCTTTCTCTGG | 720 | | |

| | | | |
|----|------|--|------|
| DB | 661 | AGATGAGGTCCTAGCTCAGGCCACATTCATTTGCTTTGTAGCTTCTTTCTGCG | 720 |
| QY | 721 | TGCTTACGGTCCCTCTCTCTGCTGCTGATGATCAAGTAGCATAGTAGTGGTTCAAA | 780 |
| DB | 721 | TGCTTACGGTCCCTCTCTCTGCTGCTGATGATCAAGTAGCATAGTAGTGGTTCAAA | 780 |
| QY | 781 | CCGGAACAGATGACCAAGCAATCAAAATAGTTTGAATCGTTTACATCTAGTACCGTCG | 840 |
| DB | 781 | CCGGAACAGATGACCAAGCAATCAAAATAGTTTGAATCGTTTACATCTAGTACCGTCG | 840 |
| QY | 841 | AACCTTACATCTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT | 900 |
| DB | 841 | AACCTTACATCTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT | 900 |
| QY | 901 | CCGATTGTCAAGTACACAAAGTACATAGTATGCGTATATGATGATGATGATGATGATGAT | 960 |
| DB | 901 | CCGATTGTCAAGTACACAAAGTACATAGTATGCGTATATGATGATGATGATGATGATGAT | 960 |
| QY | 961 | AAAGGAGGAGTACAGAGATTTTGAATCTCTGGAAGAGGATTTTCCATAGACACTA | 1020 |
| DB | 961 | AAAGGAGGAGTACAGAGATTTTGAATCTCTGGAAGAGGATTTTCCATAGACACTA | 1020 |
| QY | 1021 | ATTAGCTTTTGTGGGCGAGCTTTGTACCTATTAATGGGCTCCAAACCCCAAGTATG | 1080 |
| DB | 1021 | ATTAGCTTTTGTGGGCGAGCTTTGTACCTATTAATGGGCTCCAAACCCCAAGTATG | 1080 |
| QY | 1081 | GGCTTACAGCTTTTCCATATAAATTAAGTAATCTTTTTCCTTAAACCAATAAATAAT | 1140 |
| DB | 1081 | GGCTTACAGCTTTTCCATATAAATTAAGTAATCTTTTTCCTTAAACCAATAAATAAT | 1140 |
| QY | 1141 | ATTGAAATCTTTCCAAACCATAGAAAGTAAATTTGATCAGCGATGAAATTTTGTAC | 1200 |
| DB | 1141 | ATTGAAATCTTTCCAAACCATAGAAAGTAAATTTGATCAGCGATGAAATTTTGTAC | 1200 |
| QY | 1201 | AAAGTAGGTATTTTCAATTTGGGAGTGTACTAGTAACTAGTAACTAGTAACTAGTAACT | 1260 |
| DB | 1201 | AAAGTAGGTATTTTCAATTTGGGAGTGTACTAGTAACTAGTAACTAGTAACTAGTAACT | 1260 |
| QY | 1261 | TTCTGATTTTGGATTTTGAAGCTTTTCTTAGGTTAAATAAAGTATTTTAAACCAAT | 1320 |
| DB | 1261 | TTCTGATTTTGGATTTTGAAGCTTTTCTTAGGTTAAATAAAGTATTTTAAACCAAT | 1320 |
| QY | 1321 | AAAAGAAAACATTTTGTGAAAAGAGAAATTAAGTTTACCTGGACCCCATTTGTACAGTGG | 1380 |
| DB | 1321 | AAAAGAAAACATTTTGTGAAAAGAGAAATTAAGTTTACCTGGACCCCATTTGTACAGTGG | 1380 |
| QY | 1381 | TCCCATATATATCTGATAGAGATAGCAATAGGAAGTATTTGTTTACGCTGGTACAA | 1440 |
| DB | 1381 | TCCCATATATATCTGATAGAGATAGCAATAGGAAGTATTTGTTTACGCTGGTACAA | 1440 |
| QY | 1441 | TCGGAATGGTTTCTTTTAAAGCTCATCGAACATCAGGACCGTTGATTTTCCCGCATCAA | 1500 |
| DB | 1441 | TCGGAATGGTTTCTTTTAAAGCTCATCGAACATCAGGACCGTTGATTTTCCCGCATCAA | 1500 |
| QY | 1501 | AAAGGTTTGAATCTATCTGATGTTTCTTCTGCTCTATATATATATCTCTGACGAGTC | 1560 |
| DB | 1501 | AAAGGTTTGAATCTATCTGATGTTTCTTCTGCTCTATATATATATCTCTGACGAGTC | 1560 |
| QY | 1561 | ACATTTAGTAACTCTCTTGGACGTTAAACCGGTTAAACAGATTCTTTCCCATTTATCC | 1620 |
| DB | 1561 | ACATTTAGTAACTCTCTTGGACGTTAAACCGGTTAAACAGATTCTTTCCCATTTATCC | 1620 |
| QY | 1621 | GCTTTTAAACAACTCTCTGCTGATCTCCACCGTCCGTTTCTCTCAGCTATATTTTA | 1677 |
| DB | 1621 | GCTTTTAAACAACTCTCTGCTGATCTCCACCGTCCGTTTCTCTCAGCTATATTTTA | 1677 |

RESULT 2
 AAA88782
 ID AAA88782 standard; cDNA; 1483 BP.
 XX
 AC
 XX
 AAA88782;
 XX

19-FEB-2001 (first entry)
 Arabidopsis thaliana SSEE1 cDNA.
 SSEE1, shrunken seed gene; storage reserve; storage protein;
 oil body; transgenic plant; ss.
 Arabidopsis thaliana.
 Key Location/Qualifiers
 CDS 122..1225
 /*tag= a
 WO200061735-A1.
 19-OCT-2000.
 07-APR-2000; 2000WO-US09192.
 08-APR-1999; 99US-0128651.
 (GEO) GEN HOSPITAL CORP.
 Lin Y;
 WPI: 2000-679483/66.
 P-PSDB; ABL9718.
 Novel shrunken seed gene useful for producing transgenic plants having
 altered production of food storage reserve material, intracellular
 transport of storage protein and formation of protein or oil bodies -
 Claim 8; Page 57; 64pp; English.
 The present sequence is that of Arabidopsis thaliana SSEE1 (shrunken
 seed) cDNA, which was isolated from a seedling cDNA library using
 a partial genomic clone as probe. SSEE1 encodes a protein (see
 AAB19718) that, when expressed in a cell of a plant, modifies or
 alters the production of a food storage reserve material (e.g.
 protein, lipid or carbohydrate storage reserve), facilitates the
 intracellular transport of a storage protein, or facilitates the
 formation of protein or oil bodies. The invention provides a
 transgenic plant (or plant cell, plant tissue, plant organ or
 plant component) which includes a recombinant SSEE1 transgene that
 modifies the production of food storage reserves, thereby
 increasing nutritional value. An antisense construct is useful for
 modifying desiccation tolerance.
 Sequence 1483 BP; 461 A; 268 C; 347 G; 407 T; 0 other;
 Query Match 14.5%; Score 242.4; DB 21; Length 1483;
 Best Local Similarity 99.6%; Pred. No. 3e-46;
 Matches 243; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 2Y 1 GGTAAAGCGTTTACTATGTTTATATGATCAACGGAAGATATGCGATGTTGAATGC 60
 2b 1226 GGTAAAGCGTTTACTATGTTTATATGATCAACGGAAGATATGCGATGTTGAATGC 1285
 2Y 61 TTUUTTCAGATCATCAAGGCTCTACAGATTTCTTAGGGAATGTTTCAGGCTTTGTTA 120
 2b 1286 TTUUTTCAGATCATCAAGGCTCTACAGATTTCTTAGGGAATGTTTCAGGCTTTGTTA 1345
 2Y 121 GAAATTTGTGTTTATGCAACAGGTAGAGACATCAACATGACATGATCTGAAGAGA 180
 2b 1346 GAAATTTGTGTTTATGCAACAGGTAGAGACATCAACATGACATGATCTGAAGAGA 1405
 2Y 181 TAGCTTCTCTATGCTTAAGAAATGCGGATACGAATAAACAAGCATCAATTAAGAT 240
 2b 1406 TAGCTTCTCTATGCTTAAGAAATGCGGATACGAATAAACAAGCATCAATTAAGAT 1465
 2Y 241 TAAA 244
 2b 1466 TAAA 1469

RESULT 3
 ABZ10246
 ID ABZ10246 standard; DNA; 8056 BP.
 AC ABZ10246;
 XX
 XX 16-JAN-2003 (first entry)
 DT Haematopoietic cell proliferation disorder related DNA sequence #386.
 DE Human; haematopoietic cell proliferation disorder; cytostatic;
 XX gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
 KW cytosine methylation state; gene; ds.
 KW Homo sapiens.
 OS WO200277272-A2.
 XX
 XX 03-OCT-2002.
 PD 26-MAR-2002; 2002WO-EP03401.
 XX 26-MAR-2001; 2001US-278333P.
 XX (EPIG-) EPIGENOMICS AG.
 XX Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
 PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
 PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T;
 PI Pelet C, Schwöpe I, Ziebarth H;
 XX
 XX WPI: 2003-018942/01.
 DR Detecting and differentiating between hematopoietic cell proliferative
 XX disorders, comprises contacting a target nucleic acid with a reagent
 PT that distinguishes between methylated and non-methylated CpG
 PT dinucleotides -
 XX
 XX Claim 28; SEQ ID 386; 117pp; English.
 PS The present invention describes a method for detecting and
 CC differentiating between hematopoietic cell proliferative disorders
 CC associated with at least 1 gene and/or their regulatory regions in a
 CC subject. The method comprises contacting a target nucleic acid in a
 CC biological sample obtained from the subject with at least 1 reagent,
 CC which distinguishes between methylated and non-methylated CpG
 CC dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
 CC represent specifically claimed nucleotide sequences from the present
 CC invention. Oligonucleotides from the present invention can be used for
 CC differentiating between healthy hematopoietic cells and proliferative
 CC disorder hematopoietic cells; for differentiating between acute
 CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
 CC determining the cytosine methylation state and/or single nucleotide
 CC polymorphisms (SNPs) of hematopoietic cell proliferation disorder
 CC related sequences and their complements; and as primers for the
 CC amplification of hematopoietic cell proliferation disorder related
 CC DNA sequences. The nucleotide sequences from the present invention can
 CC also be used for detecting a predisposition to, differentiation between
 CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
 CC hematopoietic cell proliferative disorders. The present method enables
 CC a highly specific classification of hematopoietic cell proliferative
 CC disorders allowing for improved and informed treatment of patients.
 XX Sequence 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 other;
 SQ
 Query Match 3.0%; Score 50.6; DB 25; Length 8056;
 Best Local Similarity 49.4%; Pred. No. 0.088;
 Matches 131; Conservative 0; Mismatches 134; Indels 0; Gaps 0;
 1091 TTTTCCATAAAATTAAGTAAATCTTTTTCCTTACCAATAAAATTTGAAATC 1150

Db 5627 TTTTGTGTTAAATTTTAAATAATTTTATTTTTCATATAAAAAATTAATAATG 5686
Qy 1151 TTCCCAACCATAGAAAAGTTAAATTTGATCAGCGATGGAAATTTTGTACAAAGCTAGT 1210
Db 5687 ATTGAATTTGATGATAATGATTTGTGAATAATATTTGAAATTTTAAATGTTTAA 5746
Qy 1211 ATTTCAATTTGGGAGTGCTAGTAAGTAACTAGTAAGTAACTAGTAAGTAACTAGT 1270
Db 5747 TTTTGTGTTATATAAAAAATTTGATAAATGAAATTAATGAAATAATTTTATTTT 5806
Qy 1271 GGATTTTGAAGCTTTCTTCTAGGTTAAATAACAAGTATATTACTAAACAATAAAGAAA 1330
Db 5807 TTTATGTTTTTTTATTTTAAATAATTTGAAAAAATGTTATATAAAAAATTAATAAATA 5866
Qy 1331 CATTTTGTGAAAAGAGAAATPAAGT 1355
Db 5867 AAAAAATAAAAAAATTTAATTT 5891

RESULT 4

AAL15924/c

ID AAL15924 standard; cDNA; 422 BP.

XX AC AAL15924;

XX DT 07-DEC-2001 (first entry)

XX DE Human breast cancer expressed polynucleotide 8381.

XX KW Human; breast cancer; cell marker; cytostatic; ss.

XX OS Homo sapiens.

XX PN W0200151628-A2.

XX PD 19-JUL-2001.

XX PF 10-JAN-2001; 2001WO-US00798.

XX PR 14-JAN-2000; 2000US-0176077.

XX PR 14-MAR-2000; 2000US-0189167.

XX PR 24-MAR-2000; 2000US-0192099.

XX PR 29-MAR-2000; 2000US-0193480.

XX PR 15-MAY-2000; 2000US-0205230.

XX PR 09-JUN-2000; 2000US-0211315.

XX PR 25-JUL-2000; 2000US-0220534.

XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Lillie J, Xu Y, Wang Y, Steinmann K;

XX DR WPI; 2001-451856/48.

XX PT New peptide useful as a marker for the diagnosis of breast cancer -

XX PS Claim 1; Page 1516; 3695pp; English.

XX CC The invention relates to human breast cancer expressed polynucleotides

XX CC (AAL07544-AAL26789) and methods of assessing whether a patient is

XX CC afflicted with breast cancer by examining the correlation between the

XX CC expression of certain markers and the cancerous state of breast cells.

XX CC The polynucleotides and encoded polypeptides are potential markers for

XX CC detecting, diagnosing, monitoring, characterizing treating and

XX CC potentially preventing breast cancer. The polynucleotides and encoded

XX CC polypeptides are also useful for isolating compounds with cytostatic

XX CC activity.

XX CC

XX SQ Sequence 422 BP; 112 A; 32 C; 25 G; 189 T; 64 other;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Qy 1088 AGCTTTTCCATAAAATAAAGTAAATCTTTTTTGCCTAACCAATAAAATATTGAAA 1147
Db 331 ATNNNTTTGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 272
Qy 1148 ATCTTTCCCAACCATAGAAAAGTTAAATTTGATCAGCGATGGAAATTTTGTACAAAGCTA 1207
Db 271 GNNNNCANAAATTAATNTTTTNGTTTNAAAAAAANAANNTNTTTTAAATTTTTT 212
Qy 1208 GGTAATTTCAATTTGGGAGTGCTAGTAAGTAACTAGTAAGTAACTAGTAAGTAACTAGT 1267
Db 211 AAGNTTTTANNANANATTTNNNANNTTAAATTTTNNNTAAANANTAAANNCNTTTTTT 152
Qy 1268 TTTGGAATTTGAAGCTTTTCTTAGGTTAAAAACAAGTATATTACTAAACAATAAAGAA 1327
Db 151 TTTTTTTTTTATAANNNNNNNNAAAAAATAAATAAATAAATAAATAAATAAATAAATA 92
Qy 1328 AACATTTTGTGAAAAGAGAAATAAAGTTTACTGTGACCCC 1367
Db 91 AAAAAAAGAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 52

RESULT 5

AAH69544/c

ID AAH69544 standard; cDNA; 369 BP.

XX AC AAH69544;

XX DT 19-SEP-2001 (first entry)

XX DE Human cervical cancer marker nucleic acid 818.

XX KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

XX OS Homo sapiens.

XX PN W0200142467-A2.

XX PD 14-JUN-2001.

XX PF 08-DEC-2000; 2000WO-US33312.

XX PR 08-DEC-1999; 99US-0159681.

XX PR 21-DEC-1999; 99US-0171350.

XX PR 14-MAR-2000; 2000US-0189315.

XX PR 12-MAY-2000; 2000US-0203791.

XX PR 09-JUN-2000; 2000US-0210600.

XX PR 21-JUL-2000; 2000US-0220114.

XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Schlegel R, Deeds J, Berger A, Zhao X;

XX DR WPI; 2001-375006/39.

XX PT New isolated nucleic acid for diagnosing and treating cervical cancer

XX PT and for assessing and detecting compounds for treating the cancer -

XX PS Claim 1; Page 242; 1051pp; English.

XX CC The invention relates to novel genes (AAH68727-AAH73383) associated with

XX CC cervical cancer with cytostatic activity. The nucleic acids and encoded

XX CC polypeptides are useful to assess if a patient is afflicted with

XX CC cervical cancer or has a pre-malignant condition; to monitor the

XX CC progression of cervical cancer or a premalignant condition in a patient;

XX CC and to select and/or assess the efficacy of a compound or therapy for

XX CC inhibiting cervical cancer in a patient. The nucleic acids may also be

XX CC useful for gene therapy.

XX CC

XX SQ Sequence 369 BP; 96 A; 28 C; 29 G; 172 T; 44 other;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Query Match 3.0%; Score 49.6; DB 22; Length 422;

Best Local Similarity 42.1%; Pred. No. 0.072;

Matches 118; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

Query Match 2.9%; Score 49; DB 22; Length 369;

Best Local Similarity 44.4%; Pred. No. 0.096;

Matches 115; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

Y 1091 TTTTCCATAAAATTAAGTAATCTTTTGGCTAACCAATAAAATTTATTGAAATC 1150
b 313 TTGTTAAANAACCAAAAGAGTNTTTTCGGTAGGAGAAAATTTTANTTTG 254
Y 1151 TTTCCACCATAGAAAGTTAAATTTGATCAGCATGGAATTTTGTACAAAGCTAGGT 1210
b 253 ATTCATCCNNGCATGTGTTTTTANGTTTTTNCAGGCGGNNNTTTGGANAAAAAGTNA 194
Y 1211 ATTTTCATTTGGAGTGTACTAGTAACTAGTAACTAACCAAGATGAGTTTCTGATTTT 1270
b 193 NTTTTCNTTTGTTTAAATAAANAANAATTTNAAAAAANAANNNNTT 134
Y 1271 GGAATTTGAGCTTTCTTAGGTTAAAAAACAAGTATATTTACTAAACAATAAAGAAAAA 1330
b 133 NAAAAANTTAATTTTTTTTNNAAAAAANAAGTCCAAAAATNNAAANNGAAAA 74
Y 1331 CATTTTGTGAAAGAGAAA 1349
b 73 AAAAAAANAANAANAANA 55

RESULT 6
BQ67124
D ABQ67124 standard; DNA; 20933 BP.
X
C ABQ67124;
X

X 28-AUG-2002 (first entry)
E Human angiogenesis associated polynucleotide SEQ ID NO 154.
X Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
W inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcers;
W macular degeneration; inflammatory bowel disease; Crohn's disease;
W antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
W antiarteriosclerotic; ds.
X

S Homo sapiens.
N WO200246454-A2.
X
D 13-JUN-2002.
X

F 06-DEC-2001; 2001WO-EP14320.
X
X 06-DEC-2000; 2000DE-1061338.
X

X (EPIG-) EPIGENOMICS AG.
X

X Schacht O;
X

R WPI; 2002-500450/53.
X

T New nucleic acid fragments from chemically treated
T angiogenesis-associated genes, useful for determining methylation
T status, e.g. in diagnosis or treatment of cancer -
S

S Claim 1; SEQ ID NO 154; 41pp + Sequence Listing; German.
X

X The invention relates to a nucleic acid (I) comprising a segment of 19
C bases of chemically pretreated DNA of angiogenesis-associated genes (II)
C having sequences (ABQ6971-ABQ67178) or their complements. (I), also
C related oligomers, are used to evaluate the methylation status and/or
C single-nucleotide polymorphisms, in angiogenesis-related genes, for
C diagnosis and treatment of eye diseases, proliferative retinopathy,
C neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
C diabetic retinopathy, macular degeneration caused by neovascularisation,
C psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
C Crohn's disease.
C

C Note: The sequence data for this patent did not form part of the printed
C specification, but was obtained in electronic format directly from WIPO
C at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 20933 BP; 5401 A; 396 C; 5174 G; 9962 T; 0 other;
Query Match 2.9%; Score 48.6; DB 24; Length 20933;
Best Local Similarity 56.6%; Pred. No. 0.32;
Matches 90; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
QY 1091 TTTTCCATAAAATTAAGTAATCTTTTGGCTAACCAATAAAATTTATTGAAATC 1150
Db 1180 TTTTATTAATTAAGTAATTAATTTTAAAGAAAATAGGTAAAGAAATATAATGG 1239
QY 1151 TTTCCACCATAGAAAGTTAAATTTGATCAGCATGGAATTTTGTACAAAGCTAGGT 1210
Db 1240 TTTATTAAGATATATTTTATTAGATGGAATAGTTATTTTAAATATTTGTTAGTT 1299
QY 1211 ATTTTCATTTGGAGTGTACTAGTAACTAGTAACTAGTAACTAGTAA 1249
Db 1300 TTTTAAGTTGGAGTATAATAGTAAATTTTAAATTTAAA 1338

RESULT 7
ABZ12533/c
ID ABZ12533 standard; DNA; 2958 BP.
XX
AC ABZ12533;
XX

DT 21-JAN-2003 (first entry)
XX

DE Arabidopsis thaliana stress regulated gene SEQ ID NO 338.
XX

KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX

OS Arabidopsis thaliana.
XX

PN WO200216655-A2.
XX

PD 28-FEB-2002.
XX

PF 24-AUG-2001; 2001WO-US26685.
XX

PR 24-AUG-2000; 2000US-227866P.
XX

PR 26-JAN-2001; 2001US-264647P.
XX

PR 22-JUN-2001; 2001US-300111P.
XX

PA (SCRI) SCRIPPS RES INST.
XX

PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX

PI Harper JF, Kreps J, Wang X, Zhu T;
XX

DR WPI; 2002-304127/34.
XX

XX Identifying a stress condition to which a plant cell has been exposed
PT and producing plants with increased tolerance to these abiotic stresses
PT

XX Claim 144; SEQ ID NO 338; 577pp + Sequence Listing; English.
PS

XX The invention relates to identifying a stress condition to which a plant
CC cell has been exposed, comprising:
CC (a) contacting nucleic acid representative of expressed polynucleotides
CC in the plant cell with an array or probes representative of the plant
CC cell genome; and
CC (b) detecting a profile of expressed polynucleotides in the plant cell
CC characteristic of a stress response. The method is useful in the
CC production of transgenic plants, cells and seeds and in producing plants
CC with increased tolerance to abiotic stress. The present sequence is that
CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used
CC in methods of the invention.
CC

CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office.
XX

XX Sequence 2958 BP; 928 A; 621 C; 576 G; 933 T; 0 other;

Sequence 516 BP; 79 A; 14 C; 18 G; 403 T; 2 other;
XX

Query Match 2.8%; Score 46.6; DB 25; Length 516;
Best Local Similarity 46.2%; Pred. No. 0.37;
Matches 154; Conservative 0; Mismatches 179; Indels 0; Caps 0;

QY 1278 GAAGCTTTTCTTAGGTTAAAAACAAGTATATTACTAAACAATAAAAGAAAAACATTTTG 1337

Db 254 AAAAAAAAAATAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAATAA 195

1338 TGAAAAGAGAAATAAAGTTTACTGGACCCCATGTACAGATGGTCCCATATAAATACTGA 1397

Db 194 AAAAAATAAAAAAAAAAAAAAAAAATATAAAAAAAAAAAAAAAAAATAA 135

Qy 1398 TAGAATAGACCAATGGAAAGTGAATTTGTTCA 1430

Db 134 AAAAAATAAAAAAAAAAAAAATAAAATTAATACA 102

RESULTS

RESULT 9
ABK28385
TID ABK28385 STAGE 4, DNA, 7047 BB

ID ABK23385 standard; DNA; 7047 BP.
XX
NC ABK23385.

AC
 ABK28385;
 XX
 DT 22 APR 2003 (54304 054400)

DT 23-APR-2002 (first entry)
XX
BT

DE DNA transcription associated genomic DNA #130.
XX

KW DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
 KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
 KW PNA-oligonucleotide; PNA-oligo; PNA-oligo; PNA-oligo; PNA-oligo;

KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;

immunological disorder; Werner syndrome; developmental disorder;
psoriasis; Rieger's syndrome; neurological disorder; erythroplasia;
immunological disorder; Werner syndrome; developmental disorder;
psoriasis; Rieger's syndrome; neurological disorder; erythroplasia;

KW neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;

KW angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
KW polyglutamine disorder; solid tumour.

XX
OS Unidentified.

XX PN WO200192565-A2.

XX

PD 06-DEC-2001.

CX 06-APR-2001; 2001WO-EP03973.

CX 06-APR-2000; 2000DE-1019058.

CX 07-APR-2000; 2000DE-1019173.

CX 30-JUN-2000; 2000DE-1032529.

CX 01-SEP-2000; 2000DE-1043826.

CX (EPIG-) EPIGENOMICS AG.

CX Olek A, Piepenbrock C, Berlin K;

CX WPI; 2002-090046/12.

CX New nucleic acids or oligomers, useful for diagnosing or treating

CX diseases associated with DNA transcription, e.g. immunological

CX disorders, Werner syndrome, psoriasis, myocardial infarction, solid

CX tumours or cancer

CX Claim 1; SEQ ID No 259; 32pp; English.

CX The invention relates to a nucleic acid, which comprises a segment of the

CX chemically pretreated DNA of genes associated with DNA transcription from

CX one of 346 sequences, and an oligomer, in particular an oligonucleotide

CX or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical

CX to the chemically pretreated DNA of genes associated with DNA

CX transcription. The set of oligomer probes are useful for detecting the

CX cytosine methylation state and/or single nucleotide polymorphisms (SNPs)

CX in a chemically pretreated genomic DNA. The nucleic acids are useful for

CX diagnosing or treating diseases associated with DNA transcription

CX (particularly with the methylation status), e.g. adenovirus deaminase

CX deficiency, viral infection, retroviral infection, Sezary syndrome,

CX haematological disorders, immunological disorders, Werner syndrome,

CX tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,

CX neurological disorders, neurodegenerative disorders, Waardenburg

CX syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial

CX infarction, hypertension, angiogenesis, erythropoiesis, congenital heart

CX disease, HPR syndrome, arthritis, polyglutamine disorders, solid tumours

CX or cancer. Sequences ABK28127-ABK28472 represent DNA transcription

CX associated genomic DNA molecules of the invention.

CX Note: The sequence data for this patent did not form part of the printed

CX specification but was obtained in electronic format directly from the

CX European Patent Office.

CX Sequence 7047 BP; 2247 A; 75 C; 1573 G; 3151 T; 1 other;

CX Query Match 2.8%; Score 46.4; DB 24; Length 7047;

CX Best Local Similarity 49.7%; Pred. No. 0.8;

CX Matches 145; Conservative 0; Mismatches 146; Indels 1; Gaps 1;

CX 1084 TTACAGCTTTTCCATAAAATTAAGTAAATCTTTTTCCTAACCAATAAAATTTAT 1143

CX 3514 TTATATATCTATATGTTTATGTTTATGGAATGTTTATTTAGTATTTATTTAGGAAAT 3573

CX 1144 GAAATCTTTCCACATAGAAAAGTTAAATTTGATCAGCATGAAATTTTGTACAAA 1203

CX 3574 GAAATTTTCGGTTTATATA-AAAAATTTGTATACGAATGTTTATAGTATTTATTGTAA 3632

CX 1204 GCTAGTATTTTCATTTGGAGTGTACTAGTACTAGTACTAACCAATGATGTTTC 1263

CX 3633 TGGTGAATTTGTAGTAAATTTATATATTTTATTTTGGTGGTTTAAATTAATTTGGT 3692

CX 1264 TGATTTGGATTTGAAGCTTTTCTTAGGTAAAAACAAGTATATTACTAAACAATAAA 1323

CX 3693 TTATGTATTATTATGGAATTTGTTTATTTTGTATATAAAATTTGATTTATGATTTATA 3752

CX 1324 AGAAAAACATTTTGTGAAGAGAAATAAAGTTTACTGGACCCCATTTGTACA 1375

CX 3753 ATAATTTAAATAAATTTTAAAGGAATTTATGTTGAGTGAAATTAATTAATAA 3804

RESULT 10

AAS45323/c

ID AAS45323 standard; DNA; 17848 BP.

XX AC AAS45323;

XX AC AAS45323;

XX DT 18-DEC-2001 (first entry)

XX DE Chemically pretreated complementary DNA associated with cell cycle #14.

XX DE Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;

XX KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;

XX KW graft-versus-host disease; glomerular disease; Lewy body disease; cancer;

XX KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;

XX KW immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;

XX KW PCR primer.

XX KW Homo sapiens.

OS OS

XX OS

XX PN WO200168911-A2.

XX XX

XX PD 20-SEP-2001.

XX XX

XX PF 15-MAR-2001; 2001WO-EP02945.

XX PF 15-MAR-2000; 2000DE-1013847.

XX PR 06-APR-2000; 2000DE-1019058.

XX PR 07-APR-2000; 2000DE-1019173.

XX PR 30-JUN-2000; 2000DE-1032529.

XX PR 01-SEP-2000; 2000DE-1043826.

XX XX (SPIG-) EPIGENOMICS AG.

XX PA Olek A, Piepenbrock C, Berlin K;

XX PI WPI; 2001-602751/68.

XX DR

XX XX

XX PT

XX PT

XX PT

XX PT

XX XX

XX PS Claim 1; SEQ ID No 28; 28pp; English.

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX SQ

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PI Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-090046/12.
XX
XX New nucleic acids or oligomers, useful for diagnosing or treating
XX diseases associated with DNA transcription, e.g. immunological
XX disorders, Werner syndrome, psoriasis, myocardial infarction, solid
XX tumours or cancer
XX
XX Claim 1; SEQ ID No 38; 32pp; English.
XX
XX The invention relates to a nucleic acid, which comprises a segment of the
XX chemically pretreated DNA of genes associated with DNA transcription from
XX one of 345 sequences, and an oligomer, in particular an oligonucleotide
XX or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
XX to the chemically pretreated DNA of genes associated with DNA
XX transcription. The set of oligomer probes are useful for detecting the
XX cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
XX in a chemically pretreated genomic DNA. The nucleic acids are useful for
XX diagnosing or treating diseases associated with DNA transcription
XX (particularly with the methylation status), e.g. adenosine deaminase
XX deficiency, viral infection, retroviral infection, Sezary syndrome,
XX haematological disorders, immunological disorders, Werner syndrome,
XX tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
XX neurological disorders, neurodegenerative disorders, Waardenburg
XX syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
XX infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
XX disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
XX or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
XX associated genomic DNA molecules of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification but was obtained in electronic format directly from the
XX European Patent Office.
XX
XX Sequence 17848 BP; 5055 A; 211 C; 3533 G; 9043 T; 6 other;
XX
XX Query Match 2.8%; Score 46.4; DB 24; Length 17848;
XX Best Local Similarity 45.6%; Pred. No. 1;
XX Matches 164; Conservative 0; Mismatches 196; Indels 0; Gaps 0;
XX
XX 1088 AGCTTTTCCATAAATTAAAGTAATTCCTTTTGGCTTAAACATAAAATATTGAAA 1147
XX 10047 AACTTAAATAAACAATTTTAAATATTTAAATAATTTAAATTTTAAATTTCAA 9988
XX
XX 1148 ATCTTTCCACCATAGAAAGTTTAAATTTGATCAGCGATGGAAATTTTGTCACAGCTA 1207
XX 9987 AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATC 9928
XX
XX 1208 GGTATTTCATTTGGGAGTGTACTAGTAACTAGTAACTAAACCAAGATGAGTTCTGAT 1267
XX 9927 ATAAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATTT 9868
XX
XX 1268 TTGGATTGAGCTTTCTTAGGTTAAATAAACAAGTATATTTACTAACATAAAGAA 1327
XX 9867 TATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 9808
XX
XX 1328 AAACATTTTGTGAAGAGAGAAATAAAGTTTACTGGACCCCATTTGACAGATGGTCCCAT 1387
XX 9807 AATTAATAATCGAACAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 9748
XX
XX 1388 ATAACTACTGATGAGATGAGCAATGGAAGTATTTGTTCCAGCTGGTACCAATCGGAAT 1447
XX 9747 AAATCTATTACCAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 9688
XX
XX RESULT 13
XX ABZ10246/C
XX ID ABZ10246 standard; DNA; 8056 BP.
XX AC
XX ABZ10246;
XX
XX 16-JAN-2003 (first entry)
XX
DE Haematopoietic cell proliferation disorder related DNA sequence #386.
XX Human; haematopoietic cell proliferation disorder; cytostatic;
XX gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
XX cytosine methylation state; gene; ds.
XX
XX Homo sapiens.
XX
XX WO20027272-A2.
XX
XX 03-OCT-2002.
XX
XX 26-MAR-2002; 2002WO-EP03401.
XX
XX 26-MAR-2001; 2001US-278333P.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
XX Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
XX Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T;
XX Pellet C, Schwöpe I, Ziebarth R;
XX WPI; 2003-018942/01.
XX
XX Detecting and differentiating between hematopoietic cell proliferative
XX disorders, comprises contacting a target nucleic acid with a reagent
XX that distinguishes between methylated and non-methylated CpG
XX dinucleotides -
XX
XX Claim 28; SEQ ID 386; 117pp; English.
XX
XX The present invention describes a method for detecting and
XX differentiating between hematopoietic cell proliferative disorders
XX associated with at least 1 gene and/or their regulatory regions in a
XX subject. The method comprises contacting a target nucleic acid in a
XX biological sample obtained from the subject with at least 1 reagent,
XX which distinguishes between methylated and non-methylated CpG
XX dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
XX represent specifically claimed nucleotide sequences from the present
XX invention. Oligonucleotides from the present invention can be used for
XX differentiating between healthy hematopoietic cells and proliferative
XX disorder hematopoietic cells; for differentiating between acute
XX lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
XX determining the cytosine methylation state and/or single nucleotide
XX polymorphisms (SNPs) of hematopoietic cell proliferation disorder
XX related sequences and their complements; and as primers for the
XX amplification of hematopoietic cell proliferation disorder related
XX DNA sequences. The nucleotide sequences from the present invention can
XX also be used for detecting a predisposition to, differentiation between
XX subclasses, diagnosis, prognosis, treatment and/or monitoring of
XX hematopoietic cell proliferative disorders. The present method enables
XX a highly specific classification of hematopoietic cell proliferative
XX disorders allowing for improved and informed treatment of patients.
XX
XX Sequence 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 other;
XX
XX Query Match 2.7%; Score 46; DB 25; Length 8056;
XX Best Local Similarity 50.5%; Pred. No. 1;
XX Matches 112; Conservative 0; Mismatches 110; Indels 0; Gaps 0;
XX
XX 1137 AATTATTGAAAATCTTTCCAAACCATAGAAAAGTTAAATTTGATCAGCGATGGAATTTTT 1196
XX 5212 AAATATTAATTTTTTTTCAAAAATAAATAAATAAATAAATAAATAAATAAATAAATA 5153
XX
XX 1197 GTACAAAGCTAGGTATTTCAATTTGGGAGTGTACTAGTAACTAGTAACTAACAGAAAT 1256
XX 5152 TTAATTTATTTTATTTATTTATTTTAAATAAATAAATAAATAAATAAATAAATAATTT 5093
XX
XX 1257 GAGTTTCTGATTTTGGATTTTGAAGCTTTTCTTAGGTTAAAAAACAAGTATATTACTAAA 1316
XX 5092 TTTTTTTAAATTTTACATTATTATTTTTCATTTTTCATTTTTCATTTTTCATTTTTCAT 5033
XX

QY 1317 CAATAAAGAGAAAAACATTTTGTGAAAGAGAGAAATAAAGTTTA 1358
DB 5032 TCAAAATATATATATTTTATTAAATAAAAAAATAAAAAATAA 4991

RESULT 14
ABQ66998/c
ID ABQ66998 standard; DNA; 37515 BP.

AC ABQ66998;
XX
XX
DT 28-AUG-2002 (first entry)
XX
DE Human angiogenesis associated polynucleotide SEQ ID NO 28.

XX Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
KW inflammation; rheumatoid arthritis; diabetic retinopathy; antitubercles;
KW macular degeneration; inflammatory bowel disease; Crohn's disease;
KW anti-rheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiarteriosclerotic; ds.

OS Homo sapiens.
XX
XX WO200246454-A2.

XX 13-JUN-2002.

XX 06-DEC-2001; 2001WO-EP14320.

XX 06-DEC-2000; 2000DE-1061338.

XX (EPIG-) EPIGENOMICS AG.

XX Schacht O;

XX WPI; 2002-500450/53.

XX New nucleic acid fragments from chemically treated
XX angiogenesis-associated genes, useful for determining methylation
XX status, e.g. in diagnosis or treatment of cancer -

XX Claim 1; SEQ ID NO 28; 41pp + Sequence Listing; German.

XX The invention relates to a nucleic acid (I) comprising a segment of 18
XX bases of chemically pretreated DNA of angiogenesis-associated genes (II)
XX having sequences (AB666971-AB667178) or their complements (I), also
XX related oligomers, are used to evaluate the methylation status and/or
XX single-nucleotide polymorphisms, in angiogenesis-related genes, for
XX diagnosis and treatment of eye diseases, proliferative retinopathy,
XX neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
XX diabetic retinopathy, macular degeneration caused by neovascularisation,
XX psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
XX Crohn's disease.

XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 37515 BP; 11428 A; 425 C; 7165 G; 18497 T; 0 other;

Query Match 2.7%; Score 45.6; DB 24; Length 37515;
Best Local Similarity 48.6%; Pred. No. 1.8; Indels 0; Gaps 0;
Matches 123; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 1124 CCTAACCAATAAAAAATTTTGTGAAATCTTCCACACATAGAAAAGTTAAATTTGATCAGC 1183
DB 15772 CATACAAAAAAATTTACTTTAAATAATAAAAAATAAATAATCCCACTCAATATTA 15713

QY 1184 GATGGAATTTTGTACAAAGCTAGTATTTTCATTTGGAGTGCTACTAGTAACTAGTAG 1243

DB 15712 AAACAAAAATTTAAACATTTACTTAATCTAATTAATAATAATAATAATAATAATA 15653

QY 1244 TACTACCAAGATAGTTTCTGATTTTGGATTTTGGATTTTCTTAGCTTTTCTTAAAAAACA 1303

DB 15652 TACCTTTTAAATAACAATTAATACTAAAAAATAAAAAATTTCTTAAATAATAAAAA 15593
QY 1304 GTATATTACTAAACAATAAAAAAGAAAAACATTTTGTGAAAGAGAGAAATAAAGTTTACTGGA 1363
DB 15592 ACATTTTCTTAATAAAAAACAAAAACAATTTTAACCTAAAAAATAATAATAATA 15533
QY 1364 CCCATTGTACA 1375
DB 15532 ATCTTCTCTATA 15521

RESULT 15
ABL70563/c
ID ABL70563 standard; DNA; 6352 BP.

XX ABL70563;

XX 01-JUL-2002 (first entry)

XX Chemically treated cell signalling DNA sequence#227.

XX Cell signalling; cytosine methylation; cell signalling disease;
KW cancer; tumour; cytostatic; ds.

XX Unidentified.

XX WO200202807-A2.

XX 10-JAN-2002.

XX 29-JUN-2001; 2001WO-EP07471.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A. Piepenbrock C. Berlin K.

XX WPI; 2002-154758/20.

XX Nucleic acid, useful for diagnosis and therapy of diseases associated
XX with cell signalling e.g. cancer, comprises chemically modified genomic
XX sequences of genes associated with cell signalling -

XX Claim 1; SEQ ID NO 453; 24pp+sequence listing; English.

XX The invention relates to a nucleic acid comprising a sequence of at least
XX 18 bases of a segment of chemically pretreated DNA of genes associated
XX with cell signalling. The activity of the modified sequences of the
XX invention may be described as cytostatic. The object of the invention is
XX to provide the chemically modified DNA of genes associated with cell
XX signalling, as well as oligonucleotides and/or PNA-oligomers for
XX detecting cytosine methylations, as well as a method which is
XX particularly suitable for the diagnosis and/or therapy of genetic and
XX epigenetic parameters of genes associated with cell signalling. The
XX chemically modified DNA provided by the invention is useful for diagnosis
XX and therapy of diseases such as solid tumours and cancer. The sequences
XX given in records ABL70111-ABL70626 represent chemically pre-treated
XX genomic DNA's of genes associated with cell signalling.

XX Note: The sequence data for this patent is not represented in the printed
XX specification, but is based on sequence information supplied by the
XX European Patent Office.

XX Sequence 6352 BP; 1743 A; 109 C; 1264 G; 3236 T; 0 other;

Query Match 2.7%; Score 45.4; DB 24; Length 6352;
Best Local Similarity 51.4%; Pred. No. 1.3;
Matches 131; Conservative 0; Mismatches 121; Indels 3; Gaps 1;

QY 1104 TTAAGTAATCTTTTGTGCTAACCAATAAATAATTTTGAAA---ATCTTTCACCA 1160
DB 294 TTAATAAATAATCATATATCTTAACTTATATAAATAAACAACATTCATAAACAACA 235

```

2Y 1161 TAGAAAAGTTAAATTTTCATCAGCGATGGAATTTTGTACAAAGCTAGGTATTTTCATTG 1220
2b 234 AAAAAAAAAACATTAAACAATAAAAAAAGAAAAACGAAAAACCATCTATTATAATA 175
2Y 1221 GGAGTGTACTAGTAAGTACTTAACCAAGATGAGTTTCTGATTTTGGATTTTGAA 1280
2b 174 TCTCTAAACTTTATCAAACTCTATTATAAATAATTTTTTAAATTTCTCTAATATTCTAAA 115
2Y 1281 GCCTTTCTTAGGTTAAAAACAAGTATATTACTAAACAATAAAGAAAAACATTTTGTGA 1340
2b 114 CCTATCTACTAAAAAATAATTTTATTATAAACAATAAATCAAAACGTCAAA 55
2Y 1341 AAAGAGAAATAAGT 1355
2b 54 AAAAAAAAAATAATT 40

```

Search completed: February 1, 2004, 21:58:56
 Job time : 496.261 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM nucleic - nucleic search, using sw model

Run on: February 1, 2004, 21:43:46 ; Search time 3880.05 Seconds
(without alignments)
10504.671 Million cell updates/sec

Title: US-09-938-842A-3729

Perfect score: 1677

Sequence: 1 ggtaagcgttttactatg.....ttctctcagctatatattta 1677

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:

- 1: em_estba:*
- 2: em_esthm:*
- 3: em_esthm:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: em_gss_hum:*
- 18: em_gss_inv:*
- 19: em_gss_pln:*
- 20: em_gss_vrt:*
- 21: em_gss_fun:*
- 22: em_gss_man:*
- 23: em_gss_mus:*
- 24: em_gss_pro:*
- 25: em_gss_rod:*
- 26: em_gss_phg:*
- 27: em_gss_vrl:*
- 28: gb_gss1:*
- 29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 295.4 | 17.6 | 317 | 29 | BZ662927 SALK_0264 |
| 2 | 246.8 | 14.7 | 267 | 29 | AL761108 Arabidops |
| 3 | 233.2 | 13.9 | 278 | 29 | BZ377781 SALK_1061 |
| 4 | 233 | 13.9 | 347 | 9 | AV521636 AV521636 |

| | | | | | | |
|---|----|-------|-----|------|----|----------|
| C | 5 | 164.6 | 9.8 | 232 | 28 | BH612074 |
| C | 6 | 164.6 | 9.8 | 233 | 28 | BH612075 |
| | 7 | 101 | 6.0 | 486 | 9 | AV825375 |
| | 8 | 86.8 | 5.2 | 378 | 10 | BZ523004 |
| | 9 | 82 | 4.9 | 146 | 28 | BH617024 |
| C | 10 | 74.2 | 4.4 | 825 | 29 | BH493162 |
| | 11 | 64.6 | 3.9 | 735 | 29 | BZ430588 |
| C | 12 | 63.8 | 3.8 | 1201 | 13 | BX364179 |
| C | 13 | 63.4 | 3.8 | 600 | 29 | CC459772 |
| C | 14 | 62.6 | 3.7 | 1201 | 13 | BX335665 |
| C | 15 | 59.4 | 3.5 | 936 | 13 | BX404443 |
| C | 16 | 59 | 3.5 | 928 | 29 | CNS00DKY |
| | 17 | 59 | 3.5 | 999 | 13 | BX380865 |
| C | 18 | 58.8 | 3.5 | 994 | 13 | BX414650 |
| | 19 | 57.4 | 3.4 | 1101 | 29 | CNS0039G |
| C | 20 | 57.4 | 3.4 | 1101 | 29 | CNS0039Q |
| | 21 | 57 | 3.4 | 839 | 28 | AQ992711 |
| C | 22 | 57 | 3.4 | 1183 | 13 | BX417656 |
| C | 23 | 56.2 | 3.4 | 932 | 9 | AL514901 |
| | 24 | 56 | 3.3 | 771 | 13 | BX374423 |
| | 25 | 56 | 3.3 | 1201 | 13 | BX356851 |
| | 26 | 55.4 | 3.3 | 1200 | 13 | BX437758 |
| | 27 | 55 | 3.3 | 1056 | 13 | BX415058 |
| C | 28 | 55 | 3.3 | 1167 | 29 | CNS07360 |
| C | 29 | 55 | 3.3 | 1200 | 13 | BX437758 |
| C | 30 | 54.8 | 3.3 | 1101 | 29 | CNS00EPO |
| | 31 | 54.6 | 3.3 | 1064 | 13 | BX361825 |
| C | 32 | 54.6 | 3.3 | 1201 | 13 | BX399692 |
| C | 33 | 54.2 | 3.2 | 1101 | 29 | CNS0022U |
| | 34 | 54.2 | 3.2 | 1200 | 29 | CNS016CO |
| C | 35 | 54.2 | 3.2 | 1225 | 29 | CNS0161D |
| | 36 | 53.6 | 3.2 | 1133 | 13 | BX444099 |
| C | 37 | 53.4 | 3.2 | 1200 | 13 | BX447469 |
| | 38 | 53.2 | 3.2 | 1002 | 29 | CNS0163V |
| C | 39 | 53.2 | 3.2 | 1056 | 13 | BX415058 |
| | 40 | 53.2 | 3.2 | 1201 | 13 | BX420717 |
| C | 41 | 53 | 3.2 | 1101 | 29 | CNS00088 |
| | 42 | 53 | 3.2 | 1200 | 13 | BX437750 |
| C | 43 | 53 | 3.2 | 1201 | 9 | AL532464 |
| | 44 | 53 | 3.2 | 1201 | 13 | BX426629 |
| C | 45 | 52.8 | 3.1 | 1201 | 9 | AL536104 |

ALIGNMENTS

RESULT 1
BZ662927
LOCUS
DEFINITION
317 bp DNA linear GSS 31-JAN-2003
SALK_026421.45.45.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_026421.45.45.x, genomic
survey sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
BZ662927
BZ662927.1
G1:28177691
GSS
Arabidopsis thaliana (thale cress)

REFERENCE
AUTHORS
Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadriab,
, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P.,
, Zimmermann, J. and Ecker, J.R.

TITLE
A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome

JOURNAL
COMMENT
Unpublished
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379

Email: eckersalk.edu

This is single pass sequence recovered from the left border of TDNA. This sequence lies within 300 bases of the 5' end of AT2g45680.

Class: TDNA tagged.

Location/Qualifiers

1..317
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_026421.45.45.x"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html"

BASE COUNT 86 a 75 c 51 g 105 t

Query Match 17.6%; Score 295.4; DB 29; Length 317;
Best Local Similarity 99.7%; Pred. No. 1.5e-38;
Matches 296; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1391 TCCCATATATACATGATAGATAGACATGAGCATGGAAGTCTTGTTCACGTGTACAA 1440

Db 1 TCCCATATATACATGATAGATAGACATGAGCATGGAAGTCTTGTTCACGTGTACAA 60

QY 1441 TCGGAATGTTCTTTAAAGCTCATCGACACATCAGGACCGTTGATTTTCCCGCATCAA 1500

Db 61 TCGGAATGTTCTTTAAAGCTCATCGACACATCAGGACCGTTGATTTTCCCGCATCAA 120

QY 1501 AAAGCTTGATATATCTCTACCTTTTCTGCTCTATATATATATCTCTGACGATC 1560

Db 121 AAAGCTTGATATATCTCTACCTTTTCTGCTCTATATATATATCTCTGACGATC 180

QY 1561 ACATTTAGTATCTCTCTGGAGTGTACGCGTTTAAACGATTTCTTCCATTTGATCC 1620

Db 181 ACATTTAGTATCTCTCTGGAGTGTACGCGTTTAAACGATTTCTTCCATTTGATCC 240

QY 1621 GCTTTTAACTCTCGTCGTCATCCACCGTCGGTTTCTCTCAGCATATATTTTA 1677

Db 241 GCTTTTAACTCTCGTCGTCATCCACCGTCGGTTTCTCTCAGCATATATTTTA 297

RESULT 2

AL761108 267 bp DNA linear GSS 18-JUN-2002

LOCUS Arabidopsis thaliana T-DNA flanking sequence GK-205F06-014511,

DEFINITION genomic survey sequence.

ACCESSION AL761108.1 GI:21502149

VERSION GSS.

KEYWORDS Arabidopsis thaliana (thale cress)

SOURCE Arabidopsis thaliana

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 Strizhov, N., Li, Y., Rosso, M., Viehoveer, P., Dekker, K., Saedler, H.

AUTORS and Weissshaar, B.

TITLE A pipeline for automated high-throughput generation of FSTs

(flanking sequence tags) from Arabidopsis thaliana T-DNA

transformed lines

UNPUBLISHED

REFERENCE 2

ROSSO, M., Strizhov, N., Li, Y., Reiss, B., Dekker, K. and Weissshaar, B.

A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)

for flanking sequence tag based reverse genetics

UNPUBLISHED

REFERENCE 3 (bases 1 to 267)

AUTORS

Direct Submission

JOURNAL

COMMENT

Location/Qualifiers

1..267
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="GK-205F06-014511"
/clone_lib="Arabidopsis thaliana T-DNA insertion lines"
/note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector PAC161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the A. thaliana nuclear genome sequence were processed for submission. T-DNA derived sequences were removed"

BASE COUNT 64 a 48 c 68 g 87 t

ORIGIN

Query Match 14.7%; Score 246.8; DB 29; Length 267;

Best Local Similarity 99.2%; Pred. No. 1.4e-30;

Matches 248; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 385 GCAGAGTGTGGTGGCGGAAATTAACATTTATCCCTCTGTGGTGGACCGGATCTGT 444

Db 18 GCGAGAGTGTGGTGGCGGAAATTAACATTTATCCCTCTGTGGTGGACCGGATCTGT 77

QY 445 AATCGGAAAGGTGGAACCCACTTGGTTTAACTTTTAAAGCCTAAAGGTTACTACCGGTT 504

Db 78 AATCGGAAAGGTGGAACCCACTTGGTTTAACTTTTAAAGCCTAAAGGTTACTACCGGTT 137

QY 505 TGACCGGTTTAAATTTGGTGTAAATTTCTAATCCCGATCCGTTTGTGTTTAACTCTCA 564

Db 138 TGACCGGTTTAAATTTGGTGTAAATTTCTAATCCCGATCCGTTTGTGTTTAACTCTCA 197

QY 565 AGGCCACGTTATCGCCCAATATTTGATTTTGGTGGTGGGATGTTGGGTCGAATA 624

Db 198 AGGCCACGTTATCGCCCAATATTTGATTTTGGTGGTGGGATGTTGGGTCGAATA 257

QY 625 GTTGGGCGCTA 634

Db 258 GTTGGGCGCTA 267

RESULT 3

BZ377781

LOCUS

DEFINITION

SALK_106185.39.30.x Arabidopsis thaliana TDNA insertion lines

Arabidopsis thaliana genomic clone SALK_106185.39.30.x, genomic

survey sequence.

ACCESSION BZ377781

VERSION BZ377781.1 GI:25467878

KEYWORDS GSS.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 278)

Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab

, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P.

REFERENCE 2

ROSSO, M., Strizhov, N., Li, Y., Reiss, B., Dekker, K. and Weissshaar, B.

A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)

for flanking sequence tag based reverse genetics

UNPUBLISHED

REFERENCE 3

(bases 1 to 267)

TITLE Zimmerman, J. and Ecker, J.R.
A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
JOURNAL Arabidopsis Genome
COMMENT Unpublished
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of TDNA. This sequence lies within 300 bases of the 5' end of At2g45680.
Class: TDNA tagged.

FEATURES
source

Location/Qualifiers
1. .278
/organism="Arabidopsis thaliana"
/mol_type="Genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_106185.39.30.x"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/notes="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html.
72 a 69 c 44 g 93 t

BASE COUNT
ORIGIN

Query Match 13.9%; Score 233.2; DB 29; Length 278;
Best Local Similarity 96.5%; Pred. No. 2.3e-28;
Matches 249; Conservative 0; Mismatches 8; Indels 1; Gaps 1;
1421 GATTGTTCCAGTGGTACAAATCGGA-ATGGTTCCTTTAAAGCTCATCGAACACATCAGGAC 1479
1 GATTTCATCAGCTGGTACAAATCGGATGGATGGATCTTTATAGCTCATCCAAACATCAAGAG 60
1480 CGTTGATTTTCCCGCATCAAAAAGGTTGATATCTTCTCACTGTTTCTCTCTCT 1539
61 CGTTGATTTTCCCGCATCAAAAAGGTTGATATCTTCTCACTGTTTCTCTCTCT 120
1540 ATATATATATCTCGAGAGTACATTTAGTATCTCTTGGACGTGTAACGCCGTTAAAA 1599
121 ATATATATATCTCGAGAGTACATTTAGTATCTCTTGGACGTGTAACGCCGTTAAAA 180
1600 CGATTCTTCCCATTTGATCCGTTTTTAAACAATCTCGTCTCATCTCCACCGTCCGTTT 1659
181 CGATTCTTCCCATTTGATCCGTTTTTAAACAATCTCGTCTCATCTCCACCGTCCGTTT 240
1660 TCTCTCAGCTATATTTTA 1677
241 TCTCTCAGCTATATTTTA 258

RESULT 4
AV521636/c

LOCUS AV521636 347 bp mRNA linear EST 07-SEP-2000
DEFINITION AV521636 Arabidopsis thaliana aboveground organs two to six-week old Arabidopsis thaliana cDNA clone AP263b12f 3', mRNA sequence.
ACCESSION AV521636
VERSION AV521636.1 GI:86681163
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 347)
Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation

of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
20363093
PUBLISHED
10907847
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 252-0812, Japan
Email: asamizu@kazusa.or.jp, URL: <http://www.kazusa.or.jp/en/plant/>.
Location/Qualifiers
1. .347
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="AP263b12f"
/tissue_type="aboveground organs"
/dev_stage="two to six-week old"
/clone_lib="Arabidopsis thaliana aboveground organs two to six-week old"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

FEATURES
source

BASE COUNT 102 a 77 c 54 g 114 t
ORIGIN
Query Match 13.9%; Score 233; DB 9; Length 347;
Best Local Similarity 97.9%; Pred. No. 2.4e-28;
Matches 236; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 GGTAAAGCGTTTACTTATGTTTATATGCAACGGAAGATATTCATTTGTTGAATGC 60
DB 241 GGTAAAGCGTTTACTTATGTTTATATGCAACGGAAGATATTCATTTGTTGAATGC 182
QY 61 TTTTTCAGATCATCAAGGCTCTCAGATTTCTTAGGGAATGGTTTCAGGCTTTGTTA 120
DB 181 TTTTTCAGATCATCAAGGCTCTCAGATTTCTTAGGGAATGGTTTCAGGCTTTGTTA 122
QY 121 GAAATGTGTTTATTCACACAGGTAGAGAACATAACCATAGACAGATGTATCTGAAGAGA 180
DB 121 GAAATGTGTTTATTCACACAGGTAGAGACATACCATAGACAGATGTATCTGAAGAGA 62
QY 181 TAGCTTCTCTATGTCTAAAGAAATGGACCGATACGAATAAACAAGCATCATTAAAGAT 240
DB 61 TAAGCTTCTCTAGCGCTAAAGAAATGGACCGATACGAATAAACAAGCATCATTAAAGAT 2
QY 241 T 241
DB 1 T 1

RESULT 5
BH612074/c

LOCUS BH612074 232 bp DNA linear GSS 04-JAN-2002
DEFINITION SALK_032102 Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_032102, genomic survey sequence.
ACCESSION BH612074
VERSION BH612074.1 GI:18059525
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 232)
Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadriab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J. and Ecker, J.R.
A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
Unpublished
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGAL)

The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA.

Class: TDNA tagged.

Location/Qualifiers

FEATURES

source

1. .232

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/strain="Columbia 0"

/db_xref="taxon:3702"

/clone="SALK_032102"

/note="PCR was performed on Arabidopsis thaliana lines

each of which contains one or more TDNA insertion

elements. The resultant fragment for each line was

directly sequenced to determine the genomic sequence at

the site of insertion. Details of the protocols used can

be found at http://signal.salk.edu/tdna_protocols.html"

59 a 45 c 46 g 82 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 9.8%; Score 164.6; DB 28; Length 232;

Mismatches 167; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

183 AGCTTCTCTATGCTTAAGAAATGGACCGATAGCAATATAAAGCATCATTAAGATTAA 242

188 AGCTTCTCTATGCTTAAGAAATGGACCGATAGCAATATAAAGCATCATTAAGATTAA 129

243 AATGGTTTGAAGAAATACACCTTATTTATGTGAATTTGTGGTTAGTGAAGATTAA 302

128 AATGGTTTGAAGAAATACACCTTATTTATGTGAATTTGTGGTTAGTGAAGATTAA 69

303 AAACATCGGAATCCAAACCTCAAAATTTACCAATCAGCCCAAAATTATGAT 353

68 AAACATCGGAATCCAAACCTCAAAATTTACCAATCAGCCCAAAATTCGTAT 18

RESULT 6

BH612075/c

LOCUS

DEFINITION

233 bp DNA linear GSS 04-JAN-2002

SALK_032104 Arabidopsis thaliana TDNA insertion lines Arabidopsis

thaliana genomic clone SALK_032104, genomic survey sequence.

BH612075

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana (thale cress)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 233)

Alonso,J.M., Leisner,T.J., Baxarjas,P., Chen,H., Cheuk,R., Gadkinab

,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.

, Zimmerman,J. and Ecker,J.R.

A Sequence-Indexed Library of Insertion Mutations in the

Arabidopsis Genome

Unpublished

Contact: Joseph R. Ecker

Salk Institute Genomic Analysis Laboratory (SIGNAL)

The Salk Institute for Biological Studies

10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

Tel: 858 453 4100 x1752

Fax: 858 558 6379

Email: ecker@salk.edu

This is single pass sequence recovered from the left border of

TDNA.

Class: TDNA tagged.

Location/Qualifiers

TDNA.

Class: TDNA tagged.

Location/Qualifiers

TDNA.

Class: TDNA tagged.

Location/Qualifiers

TDNA.

Class: TDNA tagged.

Location/Qualifiers

TDNA.

Class: TDNA tagged.

Location/Qualifiers

TDNA.

source

1. .233

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/strain="Columbia 0"

/db_xref="taxon:3702"

/clone="SALK_032104"

/note="PCR was performed on Arabidopsis thaliana lines

each of which contains one or more TDNA insertion

elements. The resultant fragment for each line was

directly sequenced to determine the genomic sequence at

the site of insertion. Details of the protocols used can

be found at http://signal.salk.edu/tdna_protocols.html"

58 a 45 c 47 g 83 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 9.8%; Score 164.6; DB 28; Length 233;

Mismatches 167; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

183 AGCTTCTCTATGCTTAAGAAATGGACCGATAGCAATATAAAGCATCATTAAGATTAA 242

188 AGCTTCTCTATGCTTAAGAAATGGACCGATAGCAATATAAAGCATCATTAAGATTAA 129

243 AATGGTTTGAAGAAATACACCTTATTTATGTGAATTTGTGGTTAGTGAAGATTAA 302

128 AATGGTTTGAAGAAATACACCTTATTTATGTGAATTTGTGGTTAGTGAAGATTAA 69

303 AAACATCGGAATCCAAACCTCAAAATTTACCAATCAGCCCAAAATTATGAT 353

68 AAACATCGGAATCCAAACCTCAAAATTTACCAATCAGCCCAAAATTCGTAT 18

RESULT 7

AV825375

LOCUS

DEFINITION

486 bp mRNA linear EST 01-APR-2002

AV825375 RAF17 Arabidopsis thaliana cDNA clone RAFL07-08-P04 5',

mRNA sequence.

AV825375

VERSION

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana (thale cress)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 486)

Seki,M., Narusaka,M., Ishida,J., Kaniya,A., Satou,M., Nakajima,M.,

Cono,Y., Sakurai,T., Carninci,P., Kawai,J., Itoh,M., Ishii,Y.,

Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu,M., Hayashizaki,Y.

and Shinozaki,K.

Large scale analysis of Arabidopsis full-length cDNA (2002b)

Unpublished

Contact: Motoaki Seki

Plant Functional Genomics Research Group

RIKEN Genomic Sciences Center

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: 81-298-36-4359

Fax: 81-298-36-9060

Email: msekic@riken.go.jp

An Arabidopsis full-length cDNA library was constructed essentially

as reported previously (Seki et al., 1998). cDNA cleaved with BamHI

and XhoI was ligated to modified Lambda PhiC-1 vector (Garnick et

al., submitted for publication) digested with BamHI and SalI. This

clone is in a modified pBluescript vector. Please visit our web

site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further

details.

Location/Qualifiers

1. .486

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/db_xref="taxon:3702"

/clone="RAFL07-08-P04"

FEATURES

source

1. .486

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/db_xref="taxon:3702"

/clone="RAFL07-08-P04"

FEATURES

source

1. .486

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/db_xref="taxon:3702"

/clone="RAFL07-08-P04"

```

/dev stage="rosette plants"
/lab_host="Dh108"
/clone_lib="RAFL7"
/note="Site 1: BamHI; Site 2: SalI; subjected to
cold-treated (1, 2, 5, 10, 24 hr)"
BASE COUNT      131 a 118 c 117 g 117 t 3 others
ORIGIN
Query Match      6.0%; Score 101; DB 9; Length 486;
Best Local Similarity 100.0%; Pred. No. 8.2e-07;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 1577 TTGACGTGAACGCGTTAAACGATCTTTCCCATTTGATCCGTTTAAACAATCTCTGTCATCTC 1636
Db 1 TTGACGTGAACGCGTTAAACGATCTTTCCCATTTGATCCGTTTAAACAATCTCTGTCATCTC 60

2Y 1637 GTGCTATCTCCACGTCGCTTTTCTCTCAGCTATATTTTA 1677
Db 61 GTGCTATCTCCACGTCGCTTTTCTCTCAGCTATATTTTA 101

RESULT 8
LOCUS BE523004 378 bp mRNA linear EST 19-MAR-2001
DEFINITION M31B6STM Arabidopsis developing seed Arabidopsis thaliana cDNA
ACCESSION BE523004
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
REFERENCE 1 (bases 1 to 378)
AUTHORS White, J.A., Todd, J., Newman, T., Focks, N., Girke, T., Martinez de
Ilarduya, C., Jaworski, J.G., Ohlrogge, J. and Benning, C.
TITLE A new set of Arabidopsis expressed sequence tags from developing
seeds. The metabolic pathway from carbohydrates to seed oil
JOURNAL Plant Physiol. 124 (4), 1582-1594 (2000)
MEDLINE 20567808
PUBMED 11115876
COMMENT Contact: Benning, C
Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Michigan State University DNA Sequencing Facility Arabidopsis
Biological Resource Center, The Ohio State University, 309 Botany &
Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:
6142920603 TEL: 6142929371
FEATURES
source Location/Qualifiers
1..378
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="M31B6"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/clone_lib="Arabidopsis developing seed"
/note="Organ: Developing seed; Vector: pBluescript SK-;
Site 1: EcoRI; Site 2: XhoI"
BASE COUNT      108 a 97 c 81 g 92 t
ORIGIN
Query Match      5.2%; Score 86.8; DB 10; Length 378;
Best Local Similarity 97.8%; Pred. No. 0.00018;
Matches 88; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1588 ACCCGTTAAACGATCTTTCCCATTTGATCCGTTTAAACAATCTCTGTCATCTC 1647
Db 13 ACAGAGTTAAACGATCTTTCCCATTTGATCCGTTTAAACAATCTCTGTCATCTC 72

QY 1648 CACGTCGCTTTCTCTCAGCTATATTTTA 1677
Db 73 CACGTCGCTTTCTCTCAGCTATATTTTA 102

RESULT 9
LOCUS BH617024 146 bp DNA linear GSS 30-JAN-2002
DEFINITION SALK_035853 Arabidopsis thaliana TDNA insertion lines Arabidopsis
thaliana genomic clone SALK_035853, genomic survey sequence.
ACCESSION BH617024
VERSION BH617024.1 GI:18427119
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
REFERENCE 1 (bases 1 to 146)
AUTHORS Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadzinab
, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P.,
Zimmerman, J. and Ecker, J.R.
TITLE A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome
JOURNAL Unpublished
COMMENT Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA. This sequence lies within 300 bases of the 5' end of
At2g45680.
Class: TDNA tagged.
FEATURES
source Location/Qualifiers
1..146
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_035853"
/note="Arabidopsis thaliana TDNA insertion lines"
The "PCR" was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at: http://signal.salk.edu/tdna_protocols.html"
BASE COUNT      33 a 44 c 27 g 42 t
ORIGIN
Query Match      4.9%; Score 82; DB 28; Length 146;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1596 AAAACGATCTTTCCCATTTGATCCGTTTAAACAATCTCTGTCATCTC 1655
Db 1 AAAACGATCTTTCCCATTTGATCCGTTTAAACAATCTCTGTCATCTC 60

QY 1656 GTTTTCTCTCAGCTATATTTTA 1677
Db 61 GTTTTCTCTCAGCTATATTTTA 82

RESULT 10
BH493162/c

```


http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CSODL010BB12QPl&cluster=3185.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InvitroGen Corporation 1600
Faraday Avenue Genoscope sequence ID : CSODL010BB12QPl.

FEATURES
source
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODL010YC24"
/cell_type="B CELLS (RAMOS CELL LINE)" COT 25-NORMALIZED"
/cell_line="RAMOS CELL LINE"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT
25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo (dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
ASE COUNT 347 a 128 c 121 g 449 t 156 others
RIGIN
Query Match 3.8%; Score 63.8; DB 13; Length 1201;
Best Local Similarity 38.9%; Pred. No. 0.78;
Matches 144; Conservative 46; Mismatches 180; Indels 0; Gaps 0;
Y 1086 ACAGCTTTTCCATAAATAAGTAATCTTTTTCGCTAACCAATAAATAATTATGA 1145
b 832 AAAATTTTWTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 773
Y 1146 AAATCTTCCCAACCATAGAAAAGTTAAATTTGATCAGCGATGGAAATTTTGTACAAGC 1205
b 772 AAATATTTTATTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 713
Y 1206 TAGGTATTTTCATTTGGGAGTGTACTAGTAAGTACTAGTAAGTACTAGTAAGTCTG 1265
b 712 AATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 653
Y 1266 ATTTTGGATTTTGAAGCTTTCTTAGTTTAAATTTTAAATTTTAAATTTTAAAT 1325
b 652 AAAAAAATAAATTTTWTATTTAAATAAATAAATAAATAAATAAATAAATAAATA 593
Y 1326 AAAAAATTTTGTGAAAAGAGAAATAAGTTTACTGGACCCCATTTGACAGATGGTCCCA 1385
b 592 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 533
Y 1386 TAATTAATCTGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAG 1445
b 532 WTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 473
Y 1446 ATGGTCTTTT 1455
b 472 SSSSCCTTTT 463

RESULT 13
LOCUS CC459772/c
DEFINITION
Arabidopsis thaliana genomic clone SALK_133376.34.05.x, genomic
survey sequence.
ACCESSION CC459772
VERSION CC459772.1 GI:31343739
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Arabidopsi
1 (bases 1 to 600)
REFERENCE
Alonso J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab
C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P.,
Zimmerman, J. and Becker, J.R.

TITLE A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome
JOURNAL Unpublished
COMMENT Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA.

FEATURES
Location/Qualifiers
1..600
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_133376.34.05.x"
/note="PCR was performed on Arabidopsis thaliana TDNA insertion lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"
BASE COUNT 160 a 130 c 102 g 166 t 42 others
ORIGIN
Query Match 3.8%; Score 63.4; DB 29; Length 600;
Best Local Similarity 81.3%; Pred. No. 1.1;
Matches 87; Conservative 0; Mismatches 16; Indels 4; Gaps 1;
QY 1057 TAATGGGTCCCAACCCCAAGTATGGCTTACAGCTTTTCCATAAATAAAGT---AA 1112
b 475 TAATGGGTCCCAACCCCAAGTATGGCTTACAGCTTTTCTCATTAATATTAGAGTAAG 416
QY 1113 ATCTTTTGTGCTTACCAATAAATAATTATTGAAATCTTTCCAAAC 1159
b 415 ATCTTTTGTGCTTACCAATAAATAATTATTGAAATCTTTCCAAAC 369

RESULT 14
LOCUS BX335665/c
DEFINITION
BX335665 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CSODI017JU4 5-PRIME, mRNA sequence.
ACCESSION BX335665
VERSION BX335665.1 GI:30308396
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1201)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 7077.f
Contact : Feng Liang Email : fliang@lifetech.com URL :
<http://fulllength.invitrogen.com/> InvitroGen Corporation 1600
Faraday Avenue Genoscope sequence ID : CSODI017DE02QPl.

FEATURES
source
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2004, 21:45:16 ; Search time 110.457 Seconds
(without alignments)
6701.220 Million cell updates/sec

Title: US-09-938-842A-3729

Perfect score: 1677

Sequence: 1 9Gttagcgttttactatg.....tttctctcagatatatttta 1677

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents NA:*
- 1: /cgn2_6/ptcdat2/2/ina/5A COMB.seq*
 - 2: /cgn2_6/ptcdat2/2/ina/5B COMB.seq*
 - 3: /cgn2_6/ptcdat2/2/ina/6A COMB.seq*
 - 4: /cgn2_6/ptcdat2/2/ina/6B COMB.seq*
 - 5: /cgn2_6/ptcdat2/2/ina/PTCTUS COMB.seq*
 - 6: /cgn2_6/ptcdat2/2/ina/backfileseq1.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|---------|----|---------------------|
| C 1 | 46.4 | 2.8 | 7218 | 1 | US-08-332-463-14 |
| C 2 | 42.2 | 2.5 | 1666 | 1 | US-08-076-090-1 |
| C 3 | 42.2 | 2.5 | 1666 | 5 | PCT-US94-06661-1 |
| C 4 | 42.2 | 2.5 | 2500 | 1 | US-08-550-715-10 |
| C 5 | 42.2 | 2.5 | 3385 | 4 | US-09-411-449-3 |
| C 6 | 42.2 | 2.5 | 3496 | 4 | US-09-411-449-1 |
| C 7 | 42.2 | 2.5 | 3805 | 4 | US-08-411-449-4 |
| C 8 | 42.2 | 2.5 | 3916 | 4 | US-09-411-449-2 |
| C 9 | 42.2 | 2.5 | 6060 | 5 | PCT-US96-09430-7 |
| C 10 | 42 | 2.5 | 1440 | 4 | US-09-107-532A-2589 |
| C 11 | 39.6 | 2.4 | 837 | 3 | US-08-998-416-288 |
| C 12 | 39.6 | 2.4 | 7218 | 1 | US-08-232-463-14 |
| C 13 | 39 | 2.3 | 721 | 1 | US-08-451-405A-2 |
| C 14 | 39 | 2.3 | 2251 | 3 | US-08-991-677-11 |
| C 15 | 39 | 2.3 | 1664976 | 4 | US-08-916-421B-1 |
| C 16 | 39 | 2.3 | 1664976 | 4 | US-08-916-421B-1 |
| C 17 | 38.8 | 2.3 | 2169 | 4 | US-09-434-408-3 |
| C 18 | 38.6 | 2.3 | 606 | 4 | US-09-601-198-162 |
| C 19 | 38.6 | 2.3 | 1990 | 4 | US-08-961-527-232 |
| C 20 | 38.6 | 2.3 | 19124 | 2 | US-08-487-826B-13 |
| C 21 | 38.2 | 2.3 | 721 | 1 | US-08-451-405A-2 |
| C 22 | 38.2 | 2.3 | 19250 | 4 | US-08-961-527-35 |
| C 23 | 38 | 2.3 | 2341 | 3 | US-09-187-049-11 |
| C 24 | 37.8 | 2.3 | 1189 | 1 | US-08-307-591-2 |
| C 25 | 37.6 | 2.2 | 2394 | 3 | US-09-414-010-3 |
| C 26 | 37.6 | 2.2 | 8133 | 1 | US-08-480-604A-5 |
| C 27 | 37.6 | 2.2 | 8133 | 2 | US-08-405-496A-5 |

Sequence 5, Appli
Sequence 5, Appli
Sequence 5, Appli
Sequence 1, Appli
Sequence 25, Appli
Sequence 25, Appli
Sequence 25, Appli
Sequence 3, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 56, Appli
Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHREIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/232,463
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
; US-08-232-463-14

Query Match

2.8%; Score 46.4; DB 1; Length 7218;

| | | | | | |
|----|-----------------------|---|-----------------|-------------|---------|
| | Best Local Similarity | 12.9%, Pred.No. | 0.019; | Gaps | 0; |
| | Matches | S9; Conservative | 189; Mismatches | 210; Indels | 0; Gaps |
| QY | 138 | AACAGCTGAGAGACATAACCATAGACAGATGTATCTGAAGAGATAAGCTTCTCATGTCT | 197 | : | : : : : |
| Dd | 1386 | RRR | 1329 | : | : : : : |
| QY | 196 | AAAGAAATGCACCGATACGATAAACAACATCAATTAAAGATTAAATGGTTTGTAAGA | 257 | : | : : : : |
| Dd | 1328 | RRR | 1269 | : | : : : : |
| QY | 258 | ATACTACATTATTATGTGAATTGTGCTAGTGAAAAGTAAAAACAATCCGAATCCA | 317 | : | : : : : |
| Dd | 1268 | RRR | 1209 | : | : : : : |
| QY | 318 | AAACCTCAAATTACCAATCAGGCCAAATTTATGATCGTCGCATGAATGCTATGCTG | 377 | : | : : : : |
| Dd | 1208 | RRR | 1149 | : | : : : : |
| QY | 378 | ATGCTAGGCAAAAGTTGGTGCGTGCAGAAATTACAATTATCCTCTGTGTGGAGCCG | 437 | : | : : : : |
| Dd | 1148 | RRR | 1089 | : | : : : : |
| QY | 438 | AATCTGTAATCGGAAAGGTGGCAACCACTGGTTAACTTTTAAGCTTAAAGAGTTACT | 497 | : | : : : : |
| Dd | 1088 | RRRRRRRRRRRRRRRRRRRRRTCGCAAGTCCCTCGACCTCAGCACAGCTCGGAATT | 1029 | : | : : : : |
| QY | 498 | ACCGGTTTGACGGTTTATAATTGGTGTTAATTCTAATCCCGGATCCGTTTTGTTGTT | 557 | : | : : : : |
| Dd | 1028 | AATCTGTAGCGTAGGCAACAAGAGAAAATAGTTATAGTAGCGCACTCGATGGGA | 969 | : | : : : : |
| QY | 558 | AATCTCAAGCCACGTTATCGCAATATTTGATTTTT | 595 | : | : : : : |
| b | 968 | CATTTCAACCTAAACCGTTTAAATAATTTTGATCTT | 931 | : | : : : : |

RESULT 2
US-08-076-090-1
Sequence 1, Application US/08076090
Patent No. 5631162
GENERAL INFORMATION:
APPLICANT: LeBoulch, Philippe
APPLICANT: London, Irving M.
APPLICANT: Tuan, Dorothy
TITLE OF INVENTION: Retroviral Vectors for Transducing
TITLE OF INVENTION: Beta-Globulin Gene and Beta-Locus Control Region
TITLE OF INVENTION: Derivatives
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kilpatrick & Cody
STREET: 1100 Peachtree Street, Suite 2800
CITY: Atlanta
STATE: Georgia
COUNTRY: U.S.
ZIP: 30309-4530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/076,090
FILING DATE: 19930611
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: MIT 6128
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 815-6508
TELEFAX: (404) 815-6555
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

```

/ LENGTH: 1666 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ ORIGINAL SOURCE:
/ ORGANISM: Homo sapiens
/ CELL TYPE: Beta-globin gene
/ FEATURE:
/ NAME/KEY: misc signal
/ LOCATION: 37..298
/ OTHER INFORMATION: /note= "Exon III"
/ FEATURE:
/ NAME/KEY: misc_signal
/ LOCATION: 299..1148
/ OTHER INFORMATION: /note= "Intron 2"
/ FEATURE:
/ NAME/KEY: misc signal
/ LOCATION: 1149..1370
/ OTHER INFORMATION: /note= "Exon II"
/ FEATURE:
/ NAME/KEY: misc signal
/ LOCATION: 1371..1501
/ OTHER INFORMATION: /note= "Intron 1"
/ FEATURE:
/ NAME/KEY: misc signal
/ LOCATION: 1502..1643
/ OTHER INFORMATION: /note= "Exon I"
/ US-08-076-090-1
/
Query Match 2.5%; Score 42.2; DB 1; Length 1666;
Best Local Similarity 47.5%; Pred. No. 0.14;
Matches 125; Conservative 0; Mismatches 138; Indels 0; Gaps 0;
/
QY 1091 TTTTCCATAAAATTAAAGTAAATCTTTTTTTCCTAACCAATAAAAAATTATTGAAAAATC 1150
Db 694 TGTGTACACATATTAAACATTACACTTTAAACCATAAATATGATAATGATTATGTATC 743
/
QY 1151 TTTCCACCATTAGAAAGTTAAATTGTATCAGCGATGGAATTTTTGTACAAAGCTAGGT 1210
Db 744 AATTGAAATAATAAGAAATAAAGTAGGAGATTGAAATGCAATTAAGCACACATAT 803
/
QY 1211 ATTTTCATTTGGGAGTCTAGTAACTAGTAGTACTAACCAAGATCAGTTTCTGATTTT 1270
Db 804 ATTCAAATAGTAATGTACTAGCAGACTGTGTAAAGTTTTTTTTTAACTTAACTATGT 863
/
QY 1271 GGATTTTGAAGCTTTTCITTAGGTTAAAAACAGATATATTACTAAACAATAAAGAAAA 1330
Db 864 ATCTCAGAGATATTCTCTTTGTTTACACAAATGTTAAGGCATTAAAGTATAATAGTAAAA 923
/
QY 1331 CATTTTGTGAAAAAGAGAAATAAA 1353
Db 924 ATTCGGGAGAGAAAAAAGA 946
/
RESULT 3
PCT-US94-06661-1
; Sequence 1, Application PC/TUS9406661
; GENERAL INFORMATION:
/ APPLICANT:
/ TITLE OF INVENTION: Retroviral Vectors for Transducing
/ TITLE OF INVENTION: Beta-Globulin Gene and Beta-Locus Control Region
/ TITLE OF INVENTION: Derivatives
/ NUMBER OF SEQUENCES: 5
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: PCT/US94/06661

```

FILING DATE: 10-JUN-1994
CLASSIFICATION:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1666 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo Sapiens
CELL TYPE: Beta-globin gene
FEATURE:
NAME/KEY: misc signal
LOCATION: 37..298
OTHER INFORMATION: /note= "Exon III"
FEATURE:
NAME/KEY: misc signal
LOCATION: 299..1148
OTHER INFORMATION: /note= "Intron 2"
FEATURE:
NAME/KEY: misc signal
LOCATION: 1149..1370
OTHER INFORMATION: /note= "Exon II"
FEATURE:
NAME/KEY: misc signal
LOCATION: 1371..1501
OTHER INFORMATION: /note= "Intron 1"
FEATURE:
NAME/KEY: misc signal
LOCATION: 1502..1643
OTHER INFORMATION: /note= "Exon I"
CT-US94-06661-1

Query Match 2.5%; Score 42.2; DB 5; Length 1666;
Best Local Similarity 47.5%; Pred. No. 0.14;
Matches 125; Conservative 0; Mismatches 138; Indels 0; Gaps 0;
Y 1091 TTTTCCATAAATTAAGTAATCTTTTTCCTAACCAATAAAATTTGAAATC 1150
b 684 TGTGACACATATTAACACATTACACTTTAACCCATNAATATGTAATGATTATGATC 743
Y 1151 TTTCCAAACCATAGAAAGTTAAATTTGATCAGCGATGGAATTTTGTACAAAGCTAGGT 1210
b 744 AATCGAAATTAAGAAATAAAGTAGGGAGATTAATGATGCAATTAACACACATAT 803
Y 1211 ATTCATTGGGAGTGCTACTAGTAAGTAACTAACCCAGATGAGTTTCTGATTTT 1270
b 804 ATTCAAATAGTAATGCTACTAGGCAGACTGTGTAAGTTTTTTAAAGTTACTTAATGT 863
Y 1271 GGATTTTGAAGCTTTTCTTAGGTTAAACCAAGATATTAACCAATTAAGAGAAA 1330
b 864 ATCTCAGAGATATTTCCCTTTTGTATACACATGTTAAGGCATTAAATTAATAGTAAA 923
Y 1331 CATTTTGTGAAAGAGAAATAAA 1353
b 924 ATTGGGAGAGAGAAAAAAGAGA 946

RESULT 4
US-08-550-715-10/c
Sequence 10, Application US/08550715
Patent No. 5750345
GENERAL INFORMATION:
APPLICANT: Bowie, Lemuel J.
TITLE OF INVENTION: Human -Thalassemia Mutations as a Predictor of
TITLE OF INVENTION: Blood-Related Disorders
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSES: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive

CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/550,715
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gass, David A.
REGISTRATION NUMBER: 38,153
REFERENCE/DOCKET NUMBER: 28493/32834
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2500 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: CDS
LOCATION: join(687..778, 909..1131, 1982..2107)
US-08-550-715-10

Query Match 2.5%; Score 42.2; DB 1; Length 2500;
Best Local Similarity 47.5%; Pred. No. 0.16;
Matches 125; Conservative 0; Mismatches 138; Indels 0; Gaps 0;
QY 1091 TTTTCCATAAATTAAGTAATCTTTTTCCTAACCAATAAAATTTGAAATC 1150
Db 1596 TGTGACACATATTAACACATTACACTTTAACCCATNAATATGTAATGATTATGATC 1537
QY 1151 TTTCCAAACCATAGAAAGTTAAATTTGATCAGCGATGGAATTTTGTACAAAGCTAGGT 1210
Db 1536 AATTAAATTAAGAGAAATAAAGTAGGAGATTATGAATATGCAATTAAGACACATAT 1477
QY 1211 ATTTCAATTTGGGAGTGCTACTAGTAAGTAACTAACCAATGAGTTTCTGATTTT 1270
Db 1476 ATTCCAATAGTAATGCTACTAGGCAGACTGTGTAAGTTTTTTAAAGTTACTTAATGT 1417
QY 1271 GGATTTTGAAGCTTTTCTTAGGTTAAACCAAGATATTAACCAATTAAGAGAAA 1330
Db 1416 ATCTCAGAGATATTTCCCTTTTGTATACACATGTTAAGGCATTAAATTAATAGTAAA 1357
QY 1331 CATTTTGTGAAAGAGAAATAAA 1353
Db 1356 ATTGGGAGAGAGAAAAAAGAGA 1334

RESULT 5
US-09-411-449-3/c
Sequence 3, Application US/09411449
Patent No. 6524851
GENERAL INFORMATION:
APPLICANT: James Ellis
TITLE OF INVENTION: HYBRID NUCLEIC ACID MOLECULES AND VECTORS INCLUDING
TITLE OF INVENTION: beta-GLOBIN REGULATORY ELEMENTS
FILE REFERENCE: 17860017
CURRENT APPLICATION NUMBER: US/09/411,449
CURRENT FILING DATE: 1999-10-01
PRIOR APPLICATION NUMBER: 2,246,005
PRIOR FILING DATE: 1998-10-01
NUMBER OF SEQ ID NOS: 4
SOFTWARE: Patent In Ver. 2.1

```
; SEQ ID NO 3
; LENGTH: 3385
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-411-449-3

Query Match      2.5%; Score 42.2; DB 4; Length 3385;
Best Local Similarity 47.5%; Pred. No. 0.18;
Matches 125; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

QY 1091 TTTTCCATAAAATAAGTAAATCTTTTGGCTTAACCAATAAAATATTGAAAATC 1150
DB 2128 TGTGTACACATATTAAACATTCACCTTTAACCCATAAATATGATTAATGATGATC 2069
QY 1151 TTTCACCAATAGAAAAGTTAAATTTGATCAGCGATGGAATTTTGTACAAAGCTAGGT 1210
DB 2068 AATTAAAAATAAAGAAAATAAAGTAGGAGATTATGAATGCAAAATAAGCACACATAT 2009
QY 1211 ATTTCATTGGGAGTGTACTAGTAAGTAACTAAGTAACTAAGTAACTAAGTAACTAAGT 1270
DB 2008 ATTCCAAATAGTAATGTACTAGGCGAGCTGTGTAAAGTTTTTTTAAAGTTACTTAATGT 1949
QY 1271 GGATTTTGAAGCTTTCTTAGGTTAAAAAACAAGTATATTACTAAACAATAAAAGAAAAA 1330
DB 1948 ATCTCAGAGATATTTCTTTTGTATACACATGTTAAGGCAATTAAGTATAATAGTAAAA 1889
QY 1331 CATTTTGTGAAAAGAGAAATAAA 1353
DB 1888 ATTGCGGAGAGAAAAAAGAAAAGA 1866

RESULT 6
US-09-411-449-1/c
; Sequence 1, Application US/09411449
; Patent No. 6524851
; GENERAL INFORMATION:
; APPLICANT: James Ellis
; TITLE OF INVENTION: HYBRID NUCLEIC ACID MOLECULES AND VECTORS INCLUDING
; FILE REFERENCE: 17860017
; CURRENT APPLICATION NUMBER: US/09/411,449
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: 2,246,005
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 3496
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-411-449-1

Query Match      2.5%; Score 42.2; DB 4; Length 3496;
Best Local Similarity 47.5%; Pred. No. 0.18;
Matches 125; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

QY 1091 TTTTCCATAAAATAAGTAAATCTTTTGGCTTAACCAATAAAATATTGAAAATC 1150
DB 2239 TGTGTACACATATTAAACATTCACCTTTAACCCATAAATATGATTAATGATGATC 2180
QY 1151 TTTCACCAATAGAAAAGTTAAATTTGATCAGCGATGGAATTTTGTACAAAGCTAGGT 1210
DB 2179 AATTAAAAATAAAGAAAATAAAGTAGGAGATTATGAATGCAAAATAAGCACACATAT 2120
QY 1211 ATTTCATTGGGAGTGTACTAGTAAGTAACTAAGTAACTAAGTAACTAAGTAACTAAGT 1270
DB 2119 ATTCCAAATAGTAATGTACTAGGCGAGCTGTGTAAAGTTTTTTTAAAGTTACTTAATGT 2060
QY 1271 GGATTTTGAAGCTTTCTTAGGTTAAAAAACAAGTATATTACTAAACAATAAAAGAAAA 1330
DB 2059 ATCTCAGAGATATTTCTTTTGTATACACATGTTAAGGCAATTAAGTATAATAGTAAAA 2000
QY 1331 CATTTTGTGAAAAGAGAAATAAA 1353
```

```
DB 1999 ATTGCGGAGAGAAAAAAGAAAAGA 1977

RESULT 7
US-09-411-449-4/c
; Sequence 2, Application US/09411449
; Patent No. 6524851
; GENERAL INFORMATION:
; APPLICANT: James Ellis
; TITLE OF INVENTION: HYBRID NUCLEIC ACID MOLECULES AND VECTORS INCLUDING
; FILE REFERENCE: 17860017
; CURRENT APPLICATION NUMBER: US/09/411,449
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: 2,246,005
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 3805
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-411-449-4

Query Match      2.5%; Score 42.2; DB 4; Length 3805;
Best Local Similarity 47.5%; Pred. No. 0.19;
Matches 125; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

QY 1091 TTTTCCATAAAATAAGTAAATCTTTTGGCTTAACCAATAAAATATTGAAAATC 1150
DB 2548 TGTGTACACATATTAAACATTCACCTTTAACCCATAAATATGATTAATGATGATC 2489
QY 1151 TTTCACCAATAGAAAAGTTAAATTTGATCAGCGATGGAATTTTGTACAAAGCTAGGT 1210
DB 2488 AATTAAAAATAAAGAAAATAAAGTAGGAGATTATGAATGCAAAATAAGCACACATAT 2429
QY 1211 ATTTCATTGGGAGTGTACTAGTAAGTAACTAAGTAACTAAGTAACTAAGTAACTAAGT 1270
DB 2428 ATTCCAAATAGTAATGTACTAGGCGAGCTGTGTAAAGTTTTTTTAAAGTTACTTAATGT 2369
QY 1271 GGATTTTGAAGCTTTCTTAGGTTAAAAAACAAGTATATTACTAAACAATAAAAGAAAA 1330
DB 2368 ATCTCAGAGATATTTCTTTTGTATACACATGTTAAGGCAATTAAGTATAATAGTAAAA 2309
QY 1331 CATTTTGTGAAAAGAGAAATAAA 1353
DB 2308 ATTGCGGAGAGAAAAAAGAAAAGA 2286

RESULT 8
US-09-411-449-2/c
; Sequence 2, Application US/09411449
; Patent No. 6524851
; GENERAL INFORMATION:
; APPLICANT: James Ellis
; TITLE OF INVENTION: HYBRID NUCLEIC ACID MOLECULES AND VECTORS INCLUDING
; FILE REFERENCE: 17860017
; CURRENT APPLICATION NUMBER: US/09/411,449
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: 2,246,005
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 3916
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-411-449-2

Query Match      2.5%; Score 42.2; DB 4; Length 3916;
Best Local Similarity 47.5%; Pred. No. 0.19;
```

Matches 125; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

1091 TTTTCCATAAAATTAAAGTAAATCTTTTTCCTTAACCAATAAAATTTATTGAAATC 1150

2659 TGTGTACACATATTAACATTAACCTTTTAAACCAATAATATGTAATGATTATGATC 2600

1151 TTTCCAAACCATAGAAAGTTAAATTTGATCAGCGATGGAATTTTGTACAAAGCTAGGT 1210

2599 AATTAATAAATAAAGAAATAAAGTAGGAGATTAATGAATATGCAATAAAGCACACATAT 2540

1211 ATTTCATTTGGGAGTCTACTAGTAACTAGTAACTAACCAGATGAGTTCTGATTTT 1270

2539 ATTCCAAATAGTAATGATAGCGACACTGTGTAAAGTTTTTTTAAAGTTACTTTAATGT 2480

1271 GGATTTTGAAGCTTTTCTTTAGGTTAAAAACAAAGTATATTAACAACAATAAAGAAAAA 1330

2479 ATCTCAGAGATATTTCTTTTGTATACACAATGTTAAGGCATTAAAGTATATAGTAAAA 2420

1331 CATTTTGTGAAAGAGAAATATA 1353

2419 ATTGGGAGAGAAAGAAAAAAGA 2397

RESULT 9

CT-US96-09430-7/c

Sequence 7, Application PC/TUS9609430

GENERAL INFORMATION:

APPLICANT: Glazer, Peter M.

TITLE OF INVENTION: TREATMENT OF HEMOGLOBINOPATHIES

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

ADDRESSEE: OncorPharm, Inc.

STREET: 200 Perry Parkway

CITY: Gaithersburg

STATE: Maryland

COUNTRY: US

ZIP: 20877

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US96/09430

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/473,845

FILING DATE: 07-JUN-1995

ATTORNEY/AGENT INFORMATION:

NAME: Karta, Glenn E.

REGISTRATION NUMBER: 30,649

REFERENCE/DOCKET NUMBER: PA-0040

TELECOMMUNICATION INFORMATION:

TELEPHONE: 301-527-2058

TELEFAX: 301-208-6997

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 6060 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHEICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

POSITION IN GENOME:

CHROMOSOME/SEGMENT: 11

CT-US96-09430-7

Query Match 2.58; Score 42.2; DB 5; Length 6060;

Best Local Similarity 47.58; Pred. No. 0.22;

Matches 125; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

1091 TTTTCCATAAAATTAAAGTAAATCTTTTTCCTTAACCAATAAAATTTATTGAAATC 1150

3096 TGTGTACACATATTAACATTAACCTTTTAAACCAATAATATGTAATGATTATGATC 3037

1151 TTTCCAAACCATAGAAAGTTAAATTTGATCAGCGATGGAATTTTGTACAAAGCTAGGT 1210

3036 AATTAATAAATAAAGAAATAAAGTAGGAGATTAATGAATATGCAATAAAGCACACATAT 2977

1211 ATTTCATTTGGGAGTCTACTAGTAACTAGTAACTAACCAGATGAGTTCTGATTTT 1270

2976 ATTCCAAATAGTAATGATAGCGACACTGTGTAAAGTTTTTTTAAAGTTACTTTAATGT 2917

1271 GGATTTTGAAGCTTTTCTTTAGGTTAAAAACAAAGTATATTAACAACAATAAAGAAAAA 1330

2916 ATCTCAGAGATATTTCTTTTGTATACACAATGTTAAGGCATTAAAGTATATAGTAAAA 2857

1331 CATTTTGTGAAAGAGAAATATA 1353

2856 ATTGGGAGAGAAAGAAAAAAGA 2834

RESULT 10

US-09-107-532A-2589

Sequence 2589, Application US/09107532A

Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Arinello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 2589:

SEQUENCE CHARACTERISTICS:

LENGTH: 1440 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

HYPOTHEICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...1440

SEQUENCE DESCRIPTION: SEQ ID NO: 2589;
US-09-107-532A-2589

Query Match 2.5%; Score 42; DB 4; Length 1440;
Best Local Similarity 54.5%; Pred. No. 0.15;
Matches 84; Conservative 0; Mismatches 70; Indels 0; Gaps 0;
QY 1397 AAACAAGTATATTACTAACAATAAAGAAACATTTTGTGAAAGAGCAATAAAGTT 1356
DB 102 AAGAAATCTTAAACCATCAAGCAAAATTCGAAACACACAGGTGTAAAAACAGAAATCAGTT 161
QY 1357 TACTGGACCCCATTTGTACAGATGTCCTCAATAATACTGATAGAGATAGACCAATGGA 1416
DB 162 ATCTGAACAAGAATCAAGAGGCGAGTGCATCTTCTTCGAGTAGATAAAGAAATCGA 221
QY 1417 AAGTGATTTGTCTCGTGTACATTCGAATCGT 1450
DB 222 ATTGGACCGATTTCCGCGGAAAAAGTAAGCGT 255

RESULT 11
US-08-998-416-288/c
; Sequence 288, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgens
; APPLICANT: Knechtle, Philipp
; APPLICANT: Reibischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 288:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 837 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1241RP
US-08-998-416-288

Query Match 2.4%; Score 39.6; DB 3; Length 837;

Best Local Similarity 45.7%; Pred. No. 0.53;
Matches 138; Conservative 0; Mismatches 164; Indels 0; Gaps 0;
QY 1091 TTTTTCATATAAATAAAGTAAATCTTTTTTTCCTTAACCAATAAATAAATTTTGAATC 1150
DB 601 TTTATTAATTAATTAATAATGATAAATAATTAATTAATAATAATTAATTAATAATAA 542
QY 1151 TTTCCACCATAGAAAAGTTAAATTTTCATCAGCGATCGGAATTTTGTACCAAGCTAGGT 1210
DB 541 TAAATAAATAAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAA 482
QY 1211 ATTTCAATTTGGAGTGTACTAGTAACCTAGTAAGTACTTAACCAAGATGAGTTTCTGATTT 1270
DB 481 AAGTTAAATTAATTTTAATAATAATTTCTTATAAAGATTAATAATAATAATAATCAAT 422
QY 1271 GGATTTTGAAGCTTTTCTTAGGTTAAAAAACAAGTATATTTACTTAACAATAAAGAAAA 1330
DB 421 AATATTTATAAATAAGATATATATAATAAATAAATAATAATAATAATAATAATAA 362
QY 1331 CATTTTGTGAAAGAGAAATAAAGTTTACTTGGCCCATTTGTACAGATGTCCTCCATAATA 1390
DB 361 AATCTTTATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 302
QY 1391 AT 1392
DB 301 AT 300

RESULT 12
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, P.
; APPLICANT: SCHEIFLINGER, P.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

RESULT 15
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
; Patent No. 6503723
; TITLE OF INVENTION: jannaaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22

```

PRIOR APPLICATION NUMBER: US 60/024,428
PRIOR FILING DATE: 1996-08-22
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1664976
TYPE: DNA
ORGANISM: Methanococcus jannaschii
FEATURE:
NAME/KEY: misc feature
LOCATION: (28222)..(28222)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (28257)..(28258)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84773)..(84773)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84808)..(84808)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84812)..(84812)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (103998)..(103998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234514)..(234514)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature

LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (319226)..(319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559241)..(559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (600992)..(600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1130981)..(1130981)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349491)..(1349491)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

DM nucleic - nucleic search, using sw model

Run on: February 1, 2004, 21:42:55 ; Search time 630.401 Seconds
(without alignments)
9695.297 Million cell updates/sec

Title: US-09-938-842A-3729

Perfect score: 1677

Sequence: 1 GGTAAAGCGTTTACTATG.....TTTCTCTAGTATATTTA 1677

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2434939 seqs, 1822278265 residues

Total number of hits satisfying chosen parameters: 4869878

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA.*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US05_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|---------|----|----------------------|
| 1 | 1677 | 100.0 | 1677 | 10 | US-09-938-842A-3729 |
| 2 | 1677 | 100.0 | 1677 | 12 | US-09-938-842A-3729 |
| 3 | 52.4 | 3.1 | 2000 | 10 | US-09-887-576-73 |
| 4 | 49.6 | 3.0 | 858 | 15 | US-10-198-846-7035 |
| 5 | 49.2 | 2.9 | 3673778 | 13 | US-10-312-841-2 |
| 6 | 49 | 2.9 | 3673778 | 13 | US-10-312-841-1 |
| 7 | 47.8 | 2.9 | 2958 | 10 | US-09-938-842A-338 |
| 8 | 47.8 | 2.9 | 2958 | 12 | US-09-938-842A-338 |
| 9 | 46.6 | 2.8 | 516 | 10 | US-09-960-352-5785 |
| 10 | 46.4 | 2.8 | 7047 | 13 | US-10-240-453-259 |
| 11 | 46.4 | 2.8 | 17848 | 13 | US-10-240-453-38 |
| 12 | 46.4 | 2.8 | 17848 | 15 | US-10-239-676-28 |
| 13 | 45.2 | 2.7 | 616 | 13 | US-10-027-632-233686 |
| 14 | 45.2 | 2.7 | 616 | 14 | US-10-027-632-233686 |
| 15 | 45 | 2.7 | 502 | 13 | US-09-814-353-17272 |

| | | | | | | |
|------|------|-----|---------|----|---------------------|-------------------|
| C 16 | 44.8 | 2.7 | 1271 | 10 | US-09-938-842A-2829 | Sequence 2829, Ap |
| C 17 | 44.8 | 2.7 | 1271 | 12 | US-09-938-842A-2829 | Sequence 2829, Ap |
| C 18 | 44.6 | 2.7 | 556 | 13 | US-09-814-353-4676 | Sequence 4676, Ap |
| C 19 | 44.6 | 2.7 | 556 | 13 | US-09-814-353-10975 | Sequence 10975, A |
| C 20 | 44.6 | 2.7 | 2000 | 10 | US-09-938-842A-4038 | Sequence 4038, Ap |
| C 21 | 44.6 | 2.7 | 2000 | 12 | US-09-938-842A-4038 | Sequence 4038, Ap |
| C 22 | 44.2 | 2.6 | 19734 | 13 | US-10-311-455-1906 | Sequence 1906, Ap |
| C 23 | 44.2 | 2.6 | 113515 | 13 | US-10-311-455-2148 | Sequence 2148, Ap |
| C 24 | 44 | 2.6 | 7498 | 13 | US-10-311-455-230 | Sequence 230, App |
| C 25 | 43.6 | 2.6 | 11473 | 13 | US-10-311-455-1128 | Sequence 1328, Ap |
| C 26 | 43.4 | 2.6 | 6533 | 13 | US-10-240-453-257 | Sequence 257, App |
| C 27 | 43.4 | 2.6 | 640681 | 10 | US-09-790-988-1 | Sequence 1, Appl |
| C 28 | 43.2 | 2.6 | 2938 | 15 | US-10-198-846-9862 | Sequence 9862, Ap |
| C 29 | 43.2 | 2.6 | 6175 | 13 | US-10-311-455-1280 | Sequence 1280, Ap |
| C 30 | 43 | 2.6 | 6244 | 13 | US-10-311-455-458 | Sequence 458, App |
| C 31 | 43 | 2.6 | 3673778 | 13 | US-10-312-841-1 | Sequence 1, Appl |
| C 32 | 42.8 | 2.6 | 5682 | 13 | US-10-311-455-545 | Sequence 545, App |
| C 33 | 42.8 | 2.6 | 5682 | 13 | US-10-240-485-53 | Sequence 53, Appl |
| C 34 | 42.8 | 2.6 | 9180 | 13 | US-10-311-455-1938 | Sequence 1938, Ap |
| C 35 | 42.8 | 2.6 | 3673778 | 13 | US-10-312-841-2 | Sequence 2, Appl |
| C 36 | 42.6 | 2.5 | 338 | 13 | US-10-125-968-1452 | Sequence 1252, Ap |
| C 37 | 42.6 | 2.5 | 730 | 15 | US-10-198-846-2090 | Sequence 2090, Ap |
| C 38 | 42.6 | 2.5 | 17137 | 13 | US-10-311-455-164 | Sequence 164, App |
| C 39 | 42.2 | 2.5 | 6668 | 13 | US-10-311-455-1192 | Sequence 1192, Ap |
| C 40 | 42.2 | 2.5 | 7312 | 13 | US-10-311-455-1788 | Sequence 1788, Ap |
| C 41 | 42.2 | 2.5 | 8693 | 13 | US-10-311-455-1483 | Sequence 1483, Ap |
| C 42 | 42.2 | 2.5 | 8693 | 15 | US-10-172-086-37 | Sequence 37, Appl |
| C 43 | 42.2 | 2.5 | 10138 | 13 | US-10-240-453-218 | Sequence 218, App |
| C 44 | 42.2 | 2.5 | 73308 | 10 | US-09-954-456-2276 | Sequence 2276, Ap |
| C 45 | 42 | 2.5 | 1960 | 10 | US-09-938-842A-4687 | Sequence 4687, Ap |

ALIGNMENTS

RESULT 1

US-09-938-842A-3729

Sequence 3729, Application US/09938842A

Patent No. US20020160378A1

GENERAL INFORMATION:

APPLICANT: Harper, Jeff

APPLICANT: Krepis, Joel

APPLICANT: Wang, Xun

APPLICANT: Zhu, Tong

TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING

TITLE OF INVENTION: SAME AND METHODS OF USE

FILE REFERENCE: SCRIPI300-3

CURRENT APPLICATION NUMBER: US/09/938,842A

PRIOR FILING DATE: 2001-08-24

PRIOR APPLICATION NUMBER: US 60/227,866

PRIOR FILING DATE: 2000-08-24

PRIOR APPLICATION NUMBER: US 60/264,647

PRIOR FILING DATE: 2001-01-16

PRIOR APPLICATION NUMBER: US 60/300,111

PRIOR FILING DATE: 2001-06-22

NUMBER OF SEQ ID NOS: 5379

SEQ ID NO 3729

LENGTH: 1677

TYPE: DNA

ORGANISM: Arabidopsis thaliana

US-09-938-842A-3729

Query Match 100.0%; Score 1677; DB 10; Length 1677;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1677; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGTAAAGCGTTTACTATGTTTATATGCAAGCAAGATATGCAATGTTGGATGC 60

Db 1 GGTAAAGCGTTTACTATGTTTATATGCAAGCAAGATATGCAATGTTGGATGC 60

Qy 61 TTTTTCAGATCATCAAGGCTCCTACAGATTTCTTACGAGGATGTTTCAGGCTTTGTTA 120

Db 61 TTTTTCAGATCATCAAGGCTCCTACAGATTTCTTACGAGGATGTTTCAGGCTTTGTTA 120

121 GAAATTTGTTTATTTGCAACAGGTAGAGAAACATTAACCATAGACAGATGTATCTGAAGAGA 180
181 TAAGCTTCTCTATGCTAAAGAAATGGACCGATACGAATAAACAAGCATCATTAAGAT 240
181 TAAGCTTCTCTATGCTAAAGAAATGGACCGATACGAATAAACAAGCATCATTAAGAT 240
241 TAAATGGTTTGTAGAAATATACACATTTATTTATGTAAGATTTGTTGTTAGTGAAGAAT 300
241 TAAATGGTTTGTAGAAATATACACATTTATTTATGTAAGATTTGTTGTTAGTGAAGAAT 300
301 AAAACATCGGAATCCAAACCTCAAAATTTACCAATCAGCCCAAAATTTATGATGCTGGCG 360
301 AAAACATCGGAATCCAAACCTCAAAATTTACCAATCAGCCCAAAATTTATGATGCTGGCG 360
361 TAAATGAATGGTATGCTGATGTTAGCAAAAGTTGTTGCTGCGAAATTTACCAATTTATC 420
361 TAAATGAATGGTATGCTGATGTTAGCAAAAGTTGTTGCTGCGAAATTTACCAATTTATC 420
421 CCTCTGTGTGACCCGAAATCTGTAATCGGAAAGGTGGAAACCACTTGGTTTAACTTTT 480
421 CCTCTGTGTGACCCGAAATCTGTAATCGGAAAGGTGGAAACCACTTGGTTTAACTTTT 480
481 AAGCCTAAAGGTTACTACCGGTTTACCGGTTTATTAATTTGTTGTTTAACTTAACTCCC 540
481 AAGCCTAAAGGTTACTACCGGTTTACCGGTTTATTAATTTGTTGTTTAACTTAACTCCC 540
541 GGATCCGTTTGTGTTAATCTCAAGGCCAGTTATCGCCCAATTTTGAATTTTGAATG 600
541 GGATCCGTTTGTGTTAATCTCAAGGCCAGTTATCGCCCAATTTTGAATTTTGAATG 600
601 GGTAGGAATGCTGGGTCGAATAGTTGGGCTAGCCCTCAACATGTTGGACTGAAG 660
601 GGTAGGAATGCTGGGTCGAATAGTTGGGCTAGCCCTCAACATGTTGGACTGAAG 660
661 AGAGTAGGTCAGCTCAGGCCCAATTTCAATTTTCTGTTTGTAGCTTTCTTTTCTGG 720
661 AGAGTAGGTCAGCTCAGGCCCAATTTCAATTTTCTGTTTGTAGCTTTCTTTTCTGG 720
721 TGCTTACGGTCCCTTTCTGTTGCTGCTGATGTAAGTTGAACTAGTTACCGTCCG 840
721 TGCTTACGGTCCCTTTCTGTTGCTGCTGATGTAAGTTGAACTAGTTACCGTCCG 840
781 CCGGAAACAGTACCAACGAATTAAGTTGAACTAGTTACCGTCCGTTACCGTCCG 840
781 CCGGAAACAGTACCAACGAATTAAGTTGAACTAGTTACCGTCCGTTACCGTCCG 840
841 AACTTACCAATCAATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 900
841 AACTTACCAATCAATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 900
901 CCGGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 960
901 CCGGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 960
961 AAAGGACGCAATAGGAGGATTTTGGAACTCTGAAAGAGGATTTATCCATAGACACTA 1020
961 AAAGGACGCAATAGGAGGATTTTGGAACTCTGAAAGAGGATTTATCCATAGACACTA 1020
1021 ATTAGCTTTTGGTGGCGAGCTTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 1080
1021 ATTAGCTTTTGGTGGCGAGCTTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 1080
1081 GGCCTTACAGCTTTTCCATATAAATTAAGTAAATCTTTTGGCTAACCAATTAAGATTT 1140
1081 GGCCTTACAGCTTTTCCATATAAATTAAGTAAATCTTTTGGCTAACCAATTAAGATTT 1140
1141 ATTGAATTTTCCAAACCATAGAAAGTTAAATTTGATCGGATGAAATTTTGTATC 1200
1141 ATTGAATTTTCCAAACCATAGAAAGTTAAATTTGATCGGATGAAATTTTGTATC 1200
1201 AAAGCTAGTATTTCAATTTGGAGGTGTTACTAGTAACTAGTAACTAGTAACTAGTAACT 1260

1201 AAAGCTAGTATTTCAATTTGGAGGTGTTACTAGTAACTAGTAACTAGTAACTAGTAACT 1260
1261 TTCTGATTTTGGATTTGAAGCTTTTCTTAGGTTAAAAACAAGTATATTAATAACAAT 1320
1261 TTCTGATTTTGGATTTGAAGCTTTTCTTAGGTTAAAAACAAGTATATTAATAACAAT 1320
1321 AAAGGAAACAACTTTTGTGAAAGAGGAAATTAAGTTTACTGGAACCCATTTGACAGATG 1380
1321 AAAGGAAACAACTTTTGTGAAAGAGGAAATTAAGTTTACTGGAACCCATTTGACAGATG 1380
1381 TCCCATATAAATTAATCTGATAGAGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 1440
1381 TCCCATATAAATTAATCTGATAGAGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 1440
1441 TCGAATGTTTCTTTAAAGCTCATCGAAACATCAAGACCGTTGATTTTCCCGCATCAA 1500
1441 TCGAATGTTTCTTTAAAGCTCATCGAAACATCAAGACCGTTGATTTTCCCGCATCAA 1500
1501 AAAGGTTCAATTAATTTCTCACTTTGTTTCTCTCTATATATATATATATATATATATAT 1560
1501 AAAGGTTCAATTAATTTCTCACTTTGTTTCTCTCTATATATATATATATATATATATAT 1560
1561 ACATTTAGTAAATTTCTTTGGAAGTAAAGCGGTTAAAGCGATTTCTTTCCCATTTATCC 1620
1561 ACATTTAGTAAATTTCTTTGGAAGTAAAGCGGTTAAAGCGATTTCTTTCCCATTTATCC 1620
1621 GCTTTTAAACACTCTCTGTCGTCATCTCCACCGTCCGTTTCTCTCAGCTATATTTTA 1677
1621 GCTTTTAAACACTCTCTGTCGTCATCTCCACCGTCCGTTTCTCTCAGCTATATTTTA 1677

RESULT 3
US-09-887-576-73/c
; Sequence 73, Application US/09887576
; Patent No. US20020144047A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, P.
; APPLICANT: Brown, D.
; APPLICANT: Chang, H.
; APPLICANT: Zhu, T.
; APPLICANT: Han, B.
; APPLICANT: Wang, X.
; APPLICANT: Cooper, Bret
; TITLE OF INVENTION: Promoters for regulation of plant expression
; FILE REFERENCE: 1360.001US1
; CURRENT APPLICATION NUMBER: US/09/887,576
; CURRENT FILING DATE: 2001-06-25
; PRIOR APPLICATION NUMBER: US 60/213,848
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/214,087
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/258,692
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 875
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 73
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-887-576-73

Query March 3.1%; Score 52.4; DB 10; Length 2000;
Best Local Similarity 47.1%; Pred. No. 0.067;
Matches 161; Conservative 0; Mismatches 181; Indels 0; Gaps 0;
1075 AGTATGGCTTACAGCTTTTCCATATAAATTAAGTAAATCTTTTGGCTTAACCAATA 1134
1019 ATTAGGTTCTGATCTTTTAACTATAAATAAATAAATAAATAAATAAATAAATAAATAA 960
1135 AATAATTTTGAATAATCTTTCCACCATAGAAAGTTAAATTTGATCGCGATGGAATTT 1194
959 AGAATTTCAAGTACGAACAGGTCCTAAACCTGTTCTTATTTTGGTATGATATCAT 900

| | | | | | |
|------|--|------------------|---------------|-------|------|
| 1195 | TTGTCAAAAGCTAGGTATTTTCATTTGGGAGTGTA | CTAGTAAC | TAGTAAGTACTAA | CCAGA | 1254 |
| 899 | TTCAAACTTGAGAATATTTAGCATTAATACTCTACAGAAAAA | AAAAAACTTGCATTAC | 840 | | |
| 1255 | ATGAGTTTCTGATTTTGGATTTTGAAGCTTTCTTAGGTTAAAAA | CNAAGTATATTA | CTACTA | 1314 | |
| 839 | AAGTGTTTTTTTTTTTGTTTTTGAACAACGTTAAGAACCGAA | ACTTGATCAGTTATTA | 780 | | |
| 1315 | AACATAAAGAAATAAAATTTTCTGAAAAGAGAAATAAAGTTT | TCTGCA | CCCACTTGAC | 1374 | |
| 779 | GACAAAGCGAGTACCCCTTTGTGTTGTGACATCCTATAATAA | AGTTTTTTCAGGAAC | 720 | | |
| 1375 | AGATGGTCCCATATAATACTGATAGAAGATAGACAA | TGGA | 1416 | | |
| 719 | GCCTGACGGTTAATCCATCCGCTAGAAAGAAAGATCCGTCGA | 678 | | | |

```

RESULT 4
US-10-198-846-7035/c
; Sequence 7035, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7035
; LENGTH: 858
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 137..182, 184, 185, 186, 187, 188, 209, 211, 212, 217,
; LOCATION: 219, 224, 225, 236, 237, 238, 240, 241, 242, 248, 250, 252,
; LOCATION: 253, 259, 276, 279, 281, 294, 298, 304, 306, 315, 318, 319,
; LOCATION: 320, 321, 327, 338, 349, 350, 352, 355, 363, 378, 379
; OTHER INFORMATION: n = A, T, C or G
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 380..386, 389, 393, 398, 399, 400, 408, 410, 414, 434, 441,
; LOCATION: 443, 446, 453, 459, 465, 488, 490, 497, 498, 499, 501, 503,
; LOCATION: 506, 509, 513, 517, 526, 527, 528, 529, 536, 550, 557, 562,
; LOCATION: 564, 565, 573, 576, 588, 599, 604, 607, 616, 617, 619
; OTHER INFORMATION: n = A, T, C or G
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 627..628, 643, 649, 661, 668, 672, 680, 689, 698, 705, 719,
; LOCATION: 722, 735, 739, 750, 763, 765, 769, 771, 772, 774, 784, 787,
; LOCATION: 790, 791, 792, 795, 798, 805, 814, 816, 819, 820, 822, 830,
; LOCATION: 832, 833, 838, 842, 847, 849, 850, 853, 856, 857
; OTHER INFORMATION: n = A, T, C or G
US-10-198-846-7035

```

| | | | |
|----------------|------|--|------|
| D _b | 322 | GNNNCANAAATTAANTNTTTTTNGTTNAAAAAANAANNTTTTTTAAATTTTT | 263 |
| Q _y | 1208 | GGTATTTCATTTGGCAGTGCTACTAGTAACTAGTAACTAACCAAGAATGAGTTCTCAT | 1267 |
| D _b | 262 | ACGNTTTTANNANANATTTTTNNNANNNTAAATTTTTTNNTAAANANTAANNCNTTTTT | 203 |
| Q _y | 1268 | TTTGCGATTTCGAAGCTTTCTTCTTAGGTTAAAAPACAAGTATATTACTAAACAATAAAGAA | 1327 |
| D _b | 202 | TTTTTTTTTTTATAANNNNNNNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA | 143 |
| Q _y | 1328 | AAACATTTTGTGAAAGAGAAATTAAGTTTACTGGACCCC | 1367 |
| D _b | 142 | AAAAAANAAGAAAAAANAANAANAANAANAANAANAANAANAANAAGTACCTC | 103 |

```

RESULT 5
US-10-312-841-2/c
; Sequence 2, Application US/10312841
; Publication NO. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2
; LENGTH: 367378
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (379615)
; US-10-312-841-2

```

RESULT 6
US-10-312-841-1/c
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
GENERAL INFORMATION:
APPLICANT: Epigenomics AG
TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MEC
FILE REFERENCE: E01/1208/WO
CURRENT APPLICATION NUMBER: US/10/312,841
CURRENT FILING DATE: 2002-12-30
NUMBER OF SEQ ID NOS: 2

Figure 1

;; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
;; FILE REFERENCE: 16511.006/37-21(10298)C
;; CURRENT APPLICATION NUMBER: US/09/960,352
;; CURRENT FILING DATE: 2001-09-24
;; NUMBER OF SEQ ID NOS: 15112

;; SEQ ID NO 5785

;; LENGTH: 516

;; TYPE: DNA

;; ORGANISM: Bos taurus

;; FEATURE:

;; NAME/KEY: unsure

;; LOCATION: (76),(90)

;; OTHER INFORMATION: unsure at all n locations

;; OTHER INFORMATION: Clone ID: 25-LIB3057-018-Q1-K1-G1

US-09-960-352-5785

Query Match 2.8%; Score 46.6; DB 10; Length 516;

Best Local Similarity 46.2%; Pred. No. 0.78;

Matches 154; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

QY 1098 ATAAATTAAGTAAATCTTTTTCCTTAACCAATAAAATTTATTGAAATCTTTCCAA 1157

DB 434 ATAAATTAAGTAAATCTTTTTCCTTAACCAATAAAATTTATTGAAATCTTTCCAA 375

QY 1158 CCATAGAAAGTTAAATTTGTCAGCGATCGGAATTTTGTACAAAGCTAGGTATTTCTAT 1217

DB 374 TAATTAAGTAAATTTGTCAGCGATCGGAATTTTGTACAAAGCTAGGTATTTCTAT 315

QY 1218 TTGGAGCTAGTACTAGTAAAGTAAATTTGTCAGCGATCGGAATTTTGTACAAAGCTAGGTATTTCTAT 1277

DB 314 AAAAAATTAAGTAAATTTGTCAGCGATCGGAATTTTGTACAAAGCTAGGTATTTCTAT 255

QY 1278 GAAGCTTTTCTAGTAAAGTAAATTTGTCAGCGATCGGAATTTTGTACAAAGCTAGGTATTTCTAT 1337

DB 254 AAAAAATTAAGTAAATTTGTCAGCGATCGGAATTTTGTACAAAGCTAGGTATTTCTAT 195

QY 1338 TGAAGTAAAGTAAATTTGTCAGCGATCGGAATTTTGTACAAAGCTAGGTATTTCTAT 1397

DB 194 AAAAAATTAAGTAAATTTGTCAGCGATCGGAATTTTGTACAAAGCTAGGTATTTCTAT 135

QY 1398 TAGAAGTAAAGTAAATTTGTCAGCGATCGGAATTTTGTACAAAGCTAGGTATTTCTAT 1430

DB 134 AAAAAATTAAGTAAATTTGTCAGCGATCGGAATTTTGTACAAAGCTAGGTATTTCTAT 102

RESULT 10

US-10-240-453-259

;; Sequence 259, Application US/10240453

;; Publication No. US20030148326A1

;; GENERAL INFORMATION:

;; APPLICANT: OLEK, Alexander

;; APPLICANT: PIERBROCK, Christian

;; APPLICANT: BERLIN, Kurt

;; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA

;; TITLE OF INVENTION: Transcription

;; TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated

;; TITLE OF INVENTION: With DNA Transcription

;; FILE REFERENCE: 5013.1009

;; CURRENT APPLICATION NUMBER: US/10/240,453

;; CURRENT FILING DATE: 2002-10-02

;; PRIOR APPLICATION NUMBER: PCT/EP01/03973

;; PRIOR FILING DATE: 2001-04-06

;; PRIOR APPLICATION NUMBER: DE 10019058.8

;; PRIOR FILING DATE: 2000-04-06

;; PRIOR APPLICATION NUMBER: DE 10019173.8

;; PRIOR FILING DATE: 2000-04-07

;; PRIOR APPLICATION NUMBER: DE 10032529.7

;; PRIOR FILING DATE: 2000-06-30

;; PRIOR APPLICATION NUMBER: DE 10043826.1

;; PRIOR FILING DATE: 2000-09-01

;; NUMBER OF SEQ ID NOS: 350

;; SEQ ID NO 259

;; LENGTH: 7047

;; TYPE: DNA

;; ORGANISM: Artificial Sequence

;; FEATURE:

;; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

;; NAME/KEY: unsure

;; LOCATION: (182)

US-10-240-453-259

Query Match

Best Local Similarity 49.7%; Pred. No. 3.4;

Matches 145; Conservative 0; Mismatches 146; Indels 1; Gaps 1;

QY 1084 TTACAGCTTTTCCATTAAGTAAATCTTTTTCCTTAACCAATAAAATTTATTGAAATCTTTCCAA 1143

DB 3514 TTATATATGATATGTTATGTTTTCCTTAACCAATAAAATTTATTGAAATCTTTCCAA 3573

QY 1144 GAAATCTTTCCATTAAGTAAATCTTTTTCCTTAACCAATAAAATTTATTGAAATCTTTCCAA 1203

DB 3574 GAAATCTTTCCATTAAGTAAATCTTTTTCCTTAACCAATAAAATTTATTGAAATCTTTCCAA 3632

QY 1204 GCTAGGATTTTCATTTTCCTTAACCAATAAAATTTATTGAAATCTTTCCAA 1263

DB 3633 TGGTCGAAATTTTCCTTAACCAATAAAATTTATTGAAATCTTTCCAA 3692

QY 1264 TGATTTTGGATTTTCCTTAACCAATAAAATTTATTGAAATCTTTCCAA 1323

DB 3693 TTATGATTTTCCTTAACCAATAAAATTTATTGAAATCTTTCCAA 3752

QY 1324 AGAAATCAATTTTCCTTAACCAATAAAATTTATTGAAATCTTTCCAA 1375

DB 3753 ATAATTAATTAATTTTCCTTAACCAATAAAATTTATTGAAATCTTTCCAA 3804

RESULT 11

US-10-240-453-38/c

;; Sequence 38, Application US/10240453

;; Publication No. US20030148326A1

;; GENERAL INFORMATION:

;; APPLICANT: OLEK, Alexander

;; APPLICANT: PIERBROCK, Christian

;; APPLICANT: BERLIN, Kurt

;; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA

;; TITLE OF INVENTION: Transcription

;; TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated

;; TITLE OF INVENTION: With DNA Transcription

;; FILE REFERENCE: 5013.1009

;; CURRENT APPLICATION NUMBER: US/10/240,453

;; CURRENT FILING DATE: 2002-10-02

;; PRIOR APPLICATION NUMBER: PCT/EP01/03973

;; PRIOR FILING DATE: 2001-04-06

;; PRIOR APPLICATION NUMBER: DE 10019058.8

;; PRIOR FILING DATE: 2000-04-06

;; PRIOR APPLICATION NUMBER: DE 10019173.8

;; PRIOR FILING DATE: 2000-04-07

;; PRIOR APPLICATION NUMBER: DE 10032529.7

;; PRIOR FILING DATE: 2000-06-30

;; PRIOR APPLICATION NUMBER: DE 10043826.1

;; PRIOR FILING DATE: 2000-09-01

;; NUMBER OF SEQ ID NOS: 350

;; SEQ ID NO 38

;; LENGTH: 17848

;; TYPE: DNA

;; ORGANISM: Artificial Sequence

;; FEATURE:

;; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

;; NAME/KEY: unsure

;; LOCATION: (6075, 8510, 8516, 8680, 9019, 15666)

US-10-240-453-38

Query Match

Best Local Similarity 45.6%; Score 46.4; DB 13; Length 17848;

Matches 145; Conservative 0; Mismatches 146; Indels 1; Gaps 1;


```

QY 1208 GGTATTTTCATTTGGGAGTGTACTAGTACTAGTACTAACCAGAAATGAGTTTCTGAT 1267
DB 9927 ATAAAAATTAATACATAAAATTTAAAAAAAATAATAAATATTTTAAACATTTTTTAAATTTTCTTTT 9868
QY 1268 TTTGCAATTTTGAAGCTTTTCTTAGGTTTAAAAAACAAGTATATTACTAAACAATAAAAAGAA 1327
DB 9867 TATAAAAAAATTTAAACATATTTTTTAANTTAACAAAAAATAATCTAATAAAAAAATAAAAA 9808
QY 1328 AAAACATTTTGTGAAAAGAGAAATAAAGTTTACTGGACCCCAATTGTACAGATGGTCCCAT 1387
DB 9807 AATTAAATAATTCGAAACAAAAATAAAATTTACTAAAAATAATTTTTTAAACAAAAAATAAATA 9748
QY 1388 ATAATCTGTATAGAGATAGCAATGGAAGTGAATTTGTTTCAAGTGTGTACAAATCGGAAT 1447
DB 9747 AAATCTAATTAACCAATAAAAAATAATTTAATCTTAACATAAAAAACAAATTCATCCCTAAT 9688

RESULT 13
US-10-027-632-233686
; Sequence 233686, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FASTSEQ FOR WINDOWS VERSION 4.0
; SEQ ID NO 233686
; LENGTH: 616
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-233686

Query Match 2.7%; Score 45.2; DB 13; Length 616;
Best Local Similarity 62.3%; Pred. No. 1.8;
Matches 71; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 1325 GAAAAACATTTTGTGAAGAGAAATAAAGTTTACTGGACCCCAATTGTACAGATGGTCCC 1384
DB 476 GAGACTCTTTGTGTCAAAAAAATAAAGTTTATTGGACCCCAATTGTACAGATGGTCCC 535
QY 1385 ATAATACTGTATAGAGATAGCAATGGAAGTGAATTTTCAAGTGTGTAC 1438
DB 536 CTGAGATCAAGGATAGGATACAAATAAGGATAGGCTTTGSCCTCAATGAC 589

RESULT 14
US-10-027-632-233686
; Sequence 233686, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30

```

PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 233686
LENGTH: 616
TYPE: DNA
ORGANISM: Human
US-10-027-632-233686

Query Match 2.7%; Score 45.2; DB 14; Length 616;
Best Local Similarity 62.3%; Pred. No. 1.8;
Matches 71; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 1325 GAAAAACATTTGTGAAAGAGAAATAAAGTTTACTGGACCCCAATGTACAGATGTCCTCC 1384
DB 476 GAGACTCTTTGTGTCAAAAAAAGTTTATTGGACCCCAATGTACAGATGTCCTCC 535

QY 1385 ATATATATCTGATAGAGATAGACGAAAGTGAATTTGTTCACGTGGTAC 1438
DB 536 CTGAGATGAAGATAGGATACACAAATAGATAGGCCCTTGGCCCTCAATGGAC 589

RESULT 15
US-09-814-353-17272/c
Sequence 17272, Application US/09814353
Publication No. US20030165831A1
GENERAL INFORMATION:
APPLICANT: Lee, John
APPLICANT: Thompson, Pamela
APPLICANT: Lillie, James
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
IDENTIFICATION, ASSESSMENT, PREVENTION, AND
THERAPY OF OVARIAN CANCER
FILE REFERENCE: MRI-0068
CURRENT APPLICATION NUMBER: US/09/814,353
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: US 60/191,031
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: US 60/207,124
PRIOR FILING DATE: 2000-05-25
PRIOR APPLICATION NUMBER: US 60/211,940
PRIOR FILING DATE: 2000-06-15
PRIOR APPLICATION NUMBER: US 60/216,820
PRIOR FILING DATE: 2000-07-07
PRIOR APPLICATION NUMBER: US 60/220,661
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: US 60/257,672
PRIOR FILING DATE: 2000-12-21
NUMBER OF SEQ ID NOS: 22037
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 17272
LENGTH: 502
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 86
OTHER INFORMATION: n = A,T,C or G
US-09-814-353-17272

Query Match 2.7%; Score 45; DB 13; Length 502;
Best Local Similarity 49.4%; Pred. No. 1.8;
Matches 117; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

QY 1119 TTTTGCCTAACCAATAAAATTTTGAATCTTCCAAACCATAGAAAAAGTTAAATTTGA 1178
DB 347 TTTGTTTAAATTTATCTTTTATATATTTTCCCCCCCCCCCATTTTITTTTA 288

QY 1179 TCAGCGATGGAAATTTTGTACAAAGCTAGGTATTTTCATTTGGGAGTGTACTAGTAACTA 1238
DB 287 TTCCCTTTTAAATTTTGGCCCCCAACAAATTTTATTTTACACATTTTAAAAA 228

QY 1239 GTAAGTACTAACCGAATGAGTTTCTGATTTTGGATTTTGAGCTTTTCTTAGGTTAAAA 1298
DB 227 TGAATAAATAAAAAATTTTATTTTAAAAAATAAAAAAATAAAAAAATAAAAA 168

QY 1299 AACCAAGTATATTACTAAACCAATAAAAGAAAAACATTTTGTGAAAAAGAGAAATAAAGT 1355
DB 167 AAAAAAATAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAA 111

Search completed: February 2, 2004, 01:10:49
Job time : 656.401 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M nucleic - nucleic search, using sw model

un on: February 1, 2004, 21:36:50 ; Search time 4063.41 Seconds
(without alignments)
10782.626 Million cell updates/sec

title: US-09-938-842A-1034

effect score: 1071

sequence: 1 atggcgacaattcgagct.....caoggtcatcgaaacctga 1071

coring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

searched: 2888711 seqs, 20454813386 residues

total number of hits satisfying chosen parameters: 5777422

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.mu.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.other.*

33: em.htg.mus.*

34: em.htg.pln.*

35: em.htg.rtd.*

36: em.htg.mam.*

37: em.htg.vrt.*

38: em.sy.*

39: em.htgo.hum.*

40: em.htgo.mus.*

41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|-------|-------|--------|----|-----------|-------------|
| 1 | 1071 | 100.0 | 1071 | 6 | AX506339 | Sequence |
| 2 | 1071 | 100.0 | 1071 | 8 | AF370606 | Arabidops |
| 3 | 1071 | 100.0 | 1102 | 8 | AY113981 | Arabidops |
| 4 | 1071 | 100.0 | 1650 | 8 | AY056214 | Arabidops |
| 5 | 1071 | 100.0 | 91854 | 8 | AC003580 | Arabidops |
| 6 | 157.8 | 14.7 | 882 | 8 | BT008779 | Arabidops |
| 7 | 157.8 | 14.7 | 1046 | 8 | AK118579 | Arabidops |
| 8 | 157.8 | 14.7 | 1156 | 8 | AY086379 | Arabidops |
| 9 | 157.8 | 14.7 | 16203 | 8 | AB025623 | Arabidops |
| 10 | 129 | 12.0 | 1470 | 8 | AY103304 | Arabidops |
| 11 | 129 | 12.0 | 1755 | 8 | AY058874 | Arabidops |
| 12 | 129 | 12.0 | 101154 | 8 | ATF1P2 | Arabidops |
| 13 | 124.6 | 11.6 | 451 | 11 | EX322174 | Arabidops |
| 14 | 124.6 | 11.6 | 978 | 8 | AY081706 | Arabidops |
| 15 | 124.6 | 11.6 | 1440 | 8 | AY056384 | Arabidops |
| 16 | 124.6 | 11.6 | 87148 | 8 | AC013289 | Arabidops |
| 17 | 124.4 | 11.6 | 1432 | 8 | AF165924 | Gossypium |
| 18 | 121.6 | 11.4 | 1206 | 8 | BT008493 | Arabidops |
| 19 | 121.6 | 11.4 | 1479 | 8 | AY081344 | Arabidops |
| 20 | 121.6 | 11.4 | 85961 | 8 | AC079131 | Arabidops |
| 21 | 121.6 | 11.4 | 103637 | 8 | AC079604 | Arabidops |
| 22 | 118 | 11.0 | 976 | 8 | BT006043 | Arabidops |
| 23 | 118 | 11.0 | 1081 | 6 | AX027400 | Sequence |
| 24 | 118 | 11.0 | 1085 | 8 | ATH291749 | Arabidops |
| 25 | 118 | 11.0 | 1204 | 8 | AK118178 | Arabidops |
| 26 | 118 | 11.0 | 73866 | 8 | AB026649 | Arabidops |
| 27 | 117.8 | 11.0 | 1520 | 8 | DB7261 | Oryza sativ |
| 28 | 117.8 | 11.0 | 154255 | 2 | AP004632 | Oryza sat |
| 29 | 113.8 | 10.6 | 105437 | 2 | CNS07YPE | Oryza sat |
| 30 | 113 | 10.6 | 139487 | 2 | AP003513 | Oryza sat |
| 31 | 113 | 10.6 | 164978 | 2 | AP003569 | Oryza sat |
| 32 | 111.2 | 10.4 | 720 | 8 | AY072820 | Arabidops |
| 33 | 111.2 | 10.4 | 941 | 8 | AF412074 | Arabidops |
| 34 | 111.2 | 10.4 | 125411 | 8 | ATF8L15 | Arabidops |
| 35 | 110.4 | 10.3 | 74608 | 2 | AP006175 | Oryza sat |
| 36 | 110.2 | 10.3 | 155102 | 2 | AP005848 | Oryza sat |
| 37 | 108.4 | 10.1 | 95845 | 8 | AF411807 | Lycopersi |
| 38 | 105 | 9.8 | 1549 | 8 | AB071804 | Oryza sat |
| 39 | 105 | 9.8 | 193388 | 2 | AC120539 | Oryza sat |
| 40 | 104.6 | 9.8 | 138882 | 8 | AP004223 | Oryza sat |
| 41 | 104.6 | 9.8 | 146220 | 8 | AP004672 | Oryza sat |
| 42 | 102.8 | 9.6 | 1159 | 8 | AY150488 | Arabidops |
| 43 | 102.8 | 9.6 | 1508 | 8 | AY080776 | Arabidops |
| 44 | 102.8 | 9.6 | 38529 | 8 | AC069273 | Arabidops |
| 45 | 100 | 9.3 | 1304 | 8 | AK118679 | Arabidops |

ALIGNMENTS

RESULT 1

AX506339

LOCUS

DEFINITION

Accession

Version

Keywords

Source

Organism

Reference

Authors

Title

AX506339

Sequence 1034 from Patent WO0216655.

AX506339

AX506339.1 GI:23387576

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsais.

1

Harper, J.F., Kreps, J., Wang, X. and Zhu, T.

Stress-regulated genes of plants, transgenic plants containing

1071 bp

DNA

linear

PAT 27-SEP-2002

same, and methods of use
 Patent: WO 0216655-A 1034 28-FEB-2002;
 The Scripps Research Institute (US) ; Syngenta Participations AG
 (CH)

FEATURES
 Source Location/Qualifiers
 1: 1071
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /db_xref="taxon:3702"
 287 a 288 c 256 g 240 t

BASE COUNT 287 a 288 c 256 g 240 t

ORIGIN

Query Match 100.0%; Score 1071; DB 6; Length 1071;
 Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0
 Matches 1071; Conservative 0;

QY 1 ATGGCGACAATTTCAGAAGCTTTGAAGAACTTCGACGCAAAAGATCAAACTCTTAAGAGCGGTT 60
 DB 1 ATGGCGACAATTTCAGAAGCTTTGAAGAACTTCGACGCAAAAGATCAAACTCTTAAGAGCGGTT 60

QY 61 GATCTAACCATCATCAACGGCGTTCAGAAACGTCGAAACTTCAGACACCTTCCAAAGTAAAT 120
 DB 61 GATCTAACCATCATCAACGGCGTTCAGAAACGTCGAAACTTCAGACACCTTCCAAAGTAAAT 120

QY 121 CCCACAGTCAGTCTCGAGCCCAAGCGGAGCGCGTGATCGCGTCTTTTCAATGCTCTTTA 180
 DB 121 CCCACAGTCAGTCTCGAGCCCAAGCGGAGCGCGTGATCGCGTCTTTTCAATGCTCTTTA 180

QY 181 GCTCCACCGCTTTTCGACAGGACCACTTTGAAGAGAGCTTCGACTAAAGACCGTCAACG 240
 DB 181 GCTCCACCGCTTTTCGACAGGACCACTTTGAAGAGAGCTTCGACTAAAGACCGTCAACG 240

QY 241 AAGGTTGAAGGAAGGAGGAGGATACGGATGCTGCCAGCTGTCGGCTAGGATTTTT 300
 DB 241 AAGGTTGAAGGAAGGAGGAGGATACGGATGCTGCCAGCTGTCGGCTAGGATTTTT 300

QY 301 CAATTAACTCGAGAGTTAGGTCACAAATCCGACGCGGAAACGATTCGGTGGTTGTTGGAG 360
 DB 301 CAATTAACTCGAGAGTTAGGTCACAAATCCGACGCGGAAACGATTCGGTGGTTGTTGGAG 360

QY 361 AAGCGTAGCCGCGGATATAGCCGCCACGGTACGGGACGGTCCCGGCATCGCCATG 420
 DB 361 AAGCGTAGCCGCGGATATAGCCGCCACGGTACGGGACGGTCCCGGCATCGCCATG 420

QY 421 TCGGTTAAACGGAACCTTTAAAAATCCGACGACGACAAACGCTGATTCGATATGGGTGAA 480
 DB 421 TCGGTTAAACGGAACCTTTAAAAATCCGACGACGACAAACGCTGATTCGATATGGGTGAA 480

QY 481 AATCTGATCAAGAAGAAACGTTAAACGACCTTCTTAACAGTGAATATAGACATTAACGAC 540
 DB 481 AATCTGATCAAGAAGAAACGTTAAACGACCTTCTTAACAGTGAATATAGACATTAACGAC 540

QY 541 GCGGTTTCAGCTTCCTCCGGTTTACGCTCCAAATTCGACGACGACGATCCCAACTCCG 600
 DB 541 GCGGTTTCAGCTTCCTCCGGTTTACGCTCCAAATTCGACGACGACGATCCCAACTCCG 600

QY 601 CAAGCTCTGGCATCATCCACTCTGGCTCAGCAACTTTGCGCGAAGAATGTATCCGATG 660
 DB 601 CAAGCTCTGGCATCATCCACTCTGGCTCAGCAACTTTGCGCGAAGAATGTATCCGATG 660

QY 661 TGGGCTATTCCATCAAAACGATGATTCGACGCTCGGAGCTTCTCTTGATTCACAA 720
 DB 661 TGGGCTATTCCATCAAAACGATGATTCGACGCTCGGAGCTTCTCTTGATTCACAA 720

QY 721 ATCGCTGGTCCGTCGAATCAGCGCTCAGTTATAGCTTTTCCCGCGCGCGTGTTCGCG 780
 DB 721 ATCGCTGGTCCGTCGAATCAGCGCTCAGTTATAGCTTTTCCCGCGCGCGTGTTCGCG 780

QY 781 TCGTCTTAGTCGCGCTGTTCAAACAGGTTCCACGATGCGTACACCACTCTCTTTACAA 840
 DB 781 TCGTCTTAGTCGCGCTGTTCAAACAGGTTCCACGATGCGTACACCACTCTCTTTACAA 840

QY 841 GTTGTTCGAAGCAGCGGCTTTGTATCCGTTTCAGACGTTAGCGGTTCCGAATTTATCAAGA 900

LEPKAEFVMPFSMSLAPSSSTGPPLKRASTKDRHTKVGRGRRIRMPATCAARIFOL
TRLGKHSDBETIRWLENAEPALIAATGTGTVPALAMSVNGTKIPIITWADSMGE
NLWKKRSPNSSEYIDIDAVSASSLAPATITTILOPOLASSVAQQLPGMY
PWALISNAMIPTVGAFPLIPQIAPRNSOPOLLAFAAASPSVAVAQDSTWAP
PPQVPPSSGFVSVDVSNLSRATSVWAPSSSGVTGGSSSSIIATITITLDRFSL
EYKQELHOFMSTTTARSNH
1072.1102
/gene="At2g45680"

BASE COUNT 300 a 291 c 267 g 244 t
ORIGIN

Query Match 100.0%; Score 1071; DB 8; Length 1102;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1071; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|-----|--|-----|
| 2Y | 1 | ATGCGACAATTCAGAACTTGAAGAGTTCAGGCAAGATCAAACTCTAAGAGCGTT | 60 |
| DB | 1 | ATGCGACAATTCAGAACTTGAAGAGTTCAGGCAAGATCAAACTCTAAGAGCGTT | 60 |
| 2Y | 61 | GATCTAACCATCATCAACGCGCTCAGAAAGCTCGAAACTTCAAGACTTCCAAAGTAAT | 120 |
| DB | 61 | GATCTAACCATCATCAACGCGCTCAGAAAGCTCGAAACTTCAAGACTTCCAAAGTAAT | 120 |
| 2Y | 121 | CCACAGTGAAGTTCGAGCCAGGCGGAGCGGTGATCGCGTCTGTTTCAATGCTTTA | 180 |
| DB | 121 | CCACAGTGAAGTTCGAGCCAGGCGGAGCGGTGATCGCGTCTGTTTCAATGCTTTA | 180 |
| 2Y | 181 | GCTCCACCGTCTTCGACGAGCACCATTGAAGAGAGCTTCGACTAAGACCGTCAACG | 240 |
| DB | 181 | GCTCCACCGTCTTCGACGAGCACCATTGAAGAGAGCTTCGACTAAGACCGTCAACG | 240 |
| 2Y | 241 | AAGGTTGAAGAGAGGAGAGAGATACGATCGCTGCGACCGTTCGCCATCGCGATG | 300 |
| DB | 241 | AAGGTTGAAGAGAGGAGAGAGATACGATCGCTGCGACCGTTCGCCATCGCGATG | 300 |
| 2Y | 301 | CAATTAACGAGAGTGTAGTCAAAATCCGCGCGCAACAGTTCGCTGTTTGGAG | 360 |
| DB | 301 | CAATTAACGAGAGTGTAGTCAAAATCCGCGCGCAACAGTTCGCTGTTTGGAG | 360 |
| 2Y | 361 | AAGCTGAGCGCGGATTAAGCGGCAACGGTACGGGACCGGTTCCGCCATCGCGATG | 420 |
| DB | 361 | AAGCTGAGCGCGGATTAAGCGGCAACGGTACGGGACCGGTTCCGCCATCGCGATG | 420 |
| 2Y | 421 | TCGGTTAAGCGGACCTTAAATCCGCGACGACGACGACGCTGATTCGATGCGTAA | 480 |
| DB | 421 | TCGGTTAAGCGGACCTTAAATCCGCGACGACGACGACGCTGATTCGATGCGTAA | 480 |
| 2Y | 481 | AATCTGATGAAGAGAACGTAACGACCTTCTAAGCAGTGAGTATATAGACATAAGCGAC | 540 |
| DB | 481 | AATCTGATGAAGAGAACGTAACGACCTTCTAAGCAGTGAGTATATAGACATAAGCGAC | 540 |
| 2Y | 541 | CGCGTTTACGCTTCCGCTTGTAGTCCGATTCGATTCGACGACGACGATCCACCTCG | 600 |
| DB | 541 | CGCGTTTACGCTTCCGCTTGTAGTCCGATTCGATTCGACGACGACGATCCACCTCG | 600 |
| 2Y | 601 | CAAGCTCTGGCATCATCACTGTGGCTCAGCACTTCTGCGCGCAAGGATGTATCCGATG | 660 |
| DB | 601 | CAAGCTCTGGCATCATCACTGTGGCTCAGCACTTCTGCGCGCAAGGATGTATCCGATG | 660 |
| 2Y | 661 | TGGCTATTCCATCAAAAGCAATGATTCGACGCGTTCGAGCTTCTTGTATTCACAA | 720 |
| DB | 661 | TGGCTATTCCATCAAAAGCAATGATTCGACGCGTTCGAGCTTCTTGTATTCACAA | 720 |
| 2Y | 721 | ATCGCTGCTGCTGATCATGAGCTTCAGTATTAAGCTTTTCCGCGCGCTGCTCGCG | 780 |
| DB | 721 | ATCGCTGCTGCTGATCATGAGCTTCAGTATTAAGCTTTTCCGCGCGCTGCTCGCG | 780 |
| 2Y | 781 | TCGCTTTACGCTGCTGTTTCAACAGGCTTCCAGATGCTAGACCACTCTCTTACAA | 840 |
| DB | 781 | TCGCTTTACGCTGCTGTTTCAACAGGCTTCCAGATGCTAGACCACTCTCTTACAA | 840 |
| 2Y | 841 | GTTCTTCAAGCAGCGCTTGTATCGTTTTCAGAGCTTACGCGTTCGATTTATCAAG | 900 |
| DB | 841 | GTTCTTCAAGCAGCGCTTGTATCGTTTTCAGAGCTTACGCGTTCGATTTATCAAG | 900 |

FEATURES
source
1. 1660

| | | | |
|----|------|--|------|
| DB | 841 | GTGTTTCCAAAGCAGCGCTTTGTATCCGTTTTCAGAGCTTAGCGTTTCGAATTTATCAAG | 900 |
| QY | 901 | CGGAGTCGCTTATGCTCGAGCTCAAGCTCAGGCTAACAACCGGTAGTTTCATCGTCA | 960 |
| DB | 901 | CGGAGTCGCTTATGCTCGAGCTCAAGCTCAGGCTAACAACCGGTAGTTTCATCGTCA | 960 |
| QY | 961 | ATTGCAACAACAACGAGCGCACGCTGAGAGCTTCTCCTAGAGATATACGAGAAACAA | 1020 |
| DB | 961 | ATTGCAACAACAACGAGCGCACGCTGAGAGCTTCTCCTAGAGATATACGAGAAACAA | 1020 |
| QY | 1021 | GAGCTTCCAGCTTCATGACCAACCAACGAGCGCTCATCGACCACTGA | 1071 |
| DB | 1021 | GAGCTTCCAGCTTCATGACCAACCAACGAGCGCTCATCGACCACTGA | 1071 |

RESULT 4
AY056214
LOCUS
DEFINITION
Arabidopsis thaliana putative PCP2 DNA binding protein (At2g45680)
mRNA, complete cds.
ACCESSION
VERSION
AY056214.1
GI:15810350
KEYWORDS
FLI CDNA.
SOURCE
ORGANISM
Arabidopsis thaliana (thale cress)

Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1660)
2 (bases 1 to 1660)
AUTHORS
Goldsmith, A.D., Lee, J.M., Quach, H.L., Toriumi, M., Yu, G., Bowser, L.,
Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, J.,
Kamaya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B.,
Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T.,
Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K.,
Davis, R.W., Ecker, J.R., and Theologis, A.

Arabidopsis Full Length cDNA Clones
Unpublished
2 (bases 1 to 1660)
Yanada, K., Ban, J., Banno, F., Dale, J.M., Goldsmith, A.D., Lee, J.M.,
Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Yamamura, Y.,
Yu, G., Yu, S., Bowser, L., Carninci, P., Chan, H., Cheuk, R.,
Hayashizaki, Y., Ishida, J., Jones, J., Kamaya, A., Karlin-Neumann, G.,
Kawai, J., Kim, C., Koesema, E., Lam, B., Lin, J., Meyers, M.C.,
Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T.,
Satou, M., Seki, M., Shinn, P., Southwick, A., Tracy, S.E.,
Shinozaki, K., Davis, R.W., Ecker, J.R., and Theologis, A.

Direct Submission
Submitted (12-SEP-2001) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAPL cDNAs (RAPL cDNA: RIKEN
Arabidopsis Full-length cDNA): Seki, M., Narusaka, M., Ishida, J.,
Satou, M., Kamaya, A., Sakurai, T., Carninci, P., Kawai, J.,
Hayashizaki, Y., and Shinozaki, K.

The Salk, Stanford, PGECC (SSP) Consortium members carried out the
sequencing and annotation of the RAPL cDNAs: Yanada, K., Ban, J.,
Banno, F., Dale, J.M., Goldsmith, A.D., Lee, J.M., Onodera, C.S.,
Quach, H.L., Tang, C.C., Toriumi, M., Yamamura, Y., Yu, G., Yu, S.,
Bowser, L., Chen, H., Cheuk, R., Jones, J., Karlin-Neumann, G., Kim, C.,
Koesema, E., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Nguyen, M.,
Palm, C.J., Shinn, P., Southwick, A., Tracy, S.E., Davis, R.W.,
Ecker, J.R., and Theologis, A.

Yanada, K. (SSP/PGECC) and Seki, M. (RIKEN GSC) contributed equally to
this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGECC)
contributed equally to this work as PIs.

Annotation is based on the January 2002 version of the Arabidopsis
genome submitted to GenBank.
Location/Qualifiers
1. 1660


```

/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="2"
/clone="RAF107-08-P04 (R10678)"
/note="This clone is in a modified pBluescript vector
(FLC-1) as a BamHI/XhoI insert.
scotype: Columbia"
1..1660
/gene="At2g45680"
1..102
/gene="At2g45680"
103..1173
/gene="At2g45680"
/codon_start=1
/evidence=experimental
/product="putative PCF2 DNA binding protein"
/protein_id="AA107063.1"
/db_xref="GI:15810351"
/translation="WATIKLEEVAGKDOTLRVLDLTIINGVRNVESTRPQVNPVTS
LEPKAEFVPMPSFMSLAPSGTGPLKASTKDRHTKVGGRIRMPATCAARIFOL
TRELGHKSDGETIRWLENAEPALIAATGTGTVAIMASVNGTLIKIPITINADSDMG
NMKKRKRPNSEYIDISDAVSASSGLAPIATTTIOPPOLASSVVAQQLPQGY
PMWALPSNMTPTVGAFLIPQIAGPSNPOLLAPFAAASPSYVAQQAQSTARP
PPLQVPSGFSVSDVSGSLRSATSVMAPSSSGVTGSSSIATTTTTLRDFSL
EIVEKELHOFMSTTTARSSNH"
1174..1660
/gene="At2g45680"
464 a 400 c 385 g 411 t
3'UTR
ASE COUNT 464 a 400 c 385 g 411 t
RIGIN
Query Match 100.0%; Score 1071; DB 8; Length 1660;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1071; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 1 ATGGCGACAATTCAGAGCTTGAAGAGTTGCGAGCAAGATCAAACTCTAAGAGCGGTT 60
b 103 ATGGCGACAATTCAGAGCTTGAAGAGTTGCGAGCAAGATCAAACTCTAAGAGCGGTT 162
Y 61 GATCTAACCATCATCAAGCGGTGAGAACGTCGAAACTTCAAGACCTTCCAAAGTAAAT 120
b 163 GATCTAACCATCATCAAGCGGTGAGAACGTCGAAACTTCAAGACCTTCCAAAGTAAAT 222
Y 121 CCACAGTAGTCTCGAGCCCAAGCGGAGCGGTGATGCCGTGCTTTCAATGCTTTTA 180
b 223 CCACAGTAGTCTCGAGCCCAAGCGGAGCGGTGATGCCGTGCTTTCAATGCTTTTA 282
Y 181 GCTCCACCGTCTTCACAGGACCAACCATTTAGAGAGCTTCGACTAAGACCGTCAACG 240
b 283 GCTCCACCGTCTTCACAGGACCAACCATTTAGAGAGCTTCGACTAAGACCGTCAACG 342
Y 241 AAGGTTGAAGGAAGGAGGAGGATACGGATCGCTGCCAGTGTGCGGCTAGGATTTT 300
b 343 AAGGTTGAAGGAAGGAGGAGGATACGGATCGCTGCCAGTGTGCGGCTAGGATTTT 402
Y 301 CAATTAACTCGAGAGTTAGTCAAAATCCGACGCGGAAAGATTCGGTGTGTTGGAG 360
b 403 CAATTAACTCGAGAGTTAGTCAAAATCCGACGCGGAAAGATTCGGTGTGTTGGAG 462
Y 361 AAGCGTGAGCGCGGATATATAGCCGACGCGGTACGGAAAGGTTCCGCGCATCGCCATG 420
b 463 AAGCGTGAGCGCGGATATATAGCCGACGCGGTACGGAAAGGTTCCGCGCATCGCCATG 522
Y 421 TCGGTTAAGCGAACCTTAAAAATCCCGACGACGACGACGACGACGACGACGACGACG 480
b 523 TCGGTTAAGCGAACCTTAAAAATCCCGACGACGACGACGACGACGACGACGACGACG 582
Y 481 AATCTGATGAAGAAAGAGTAAACGACCTTCTAAGTGTATATAGACATAGCGAC 540
b 583 AATCTGATGAAGAAAGAGTAAACGACCTTCTAAGTGTATATAGACATAGCGAC 642
Y 541 GCGGTTTCAGCTTCTCCGGTTTACCTCAATTTGCGAGGACGACGACGACGACGACG 600

```

```

Db 643 GCGGTTTCAGCTTCTCCCGTTTAGCTCAANTGCGACGACGACGATCAACACTCG 702
Qy 601 CAAAGCTCTGCGCATCATCCACTGTGCTCAGCAACTTCTCGCGCAAGGATGATCCGATG 660
Db 703 CAAAGCTCTGCGCATCATCCACTGTGCTCAGCAACTTCTCGCGCAAGGATGATCCGATG 762
Qy 661 TGGGCTATTCCATCAACGCAATGATTCGAGCGGTGCGAGCTTCTTCTGATTCACAA 720
Db 763 TGGGCTATTCCATCAACGCAATGATTCGAGCGGTGCGAGCTTCTTCTGATTCACAA 822
Qy 721 ATGCTGTGTCGCTCGAATCAGCTCAGTTTATAGCTTTTCCGCGCGCGCTGCTTCGCG 780
Db 823 ATGCTGTGTCGCTCGAATCAGCTCAGTTTATAGCTTTTCCGCGCGCGCTGCTTCGCG 882
Qy 781 TCGTCTTAGCTGCGCGCTGTTCAACAGGCTTCCAGATGCTAGACCACTCTTTTACAA 840
Db 883 TCGTCTTAGCTGCGCGCTGTTCAACAGGCTTCCAGATGCTAGACCACTCTTTTACAA 942
Qy 841 GTTGTTCACAGCAGCGGCTTGTATCCGTTTCAGACGTTAGCGGTTTCAATTTTCAAGA 900
Db 943 GTTGTTCACAGCAGCGGCTTGTATCCGTTTCAGACGTTAGCGGTTTCAATTTTCAAGA 1002
Qy 901 GCGACTCGGTTATGCTCCGAGCTCAAGCTCAGCGGTAAACCGGTAGTTCATCGTCA 960
Db 1003 GCGACTCGGTTATGCTCCGAGCTCAAGCTCAGCGGTAAACCGGTAGTTCATCGTCA 1062
Qy 961 ATTGCAACACAGCAGCGGCTGAGAGCTTCTCCCTAGAGATATACGAGAAACAA 1020
Db 1063 ATTGCAACACAGCAGCGGCTGAGAGCTTCTCCCTAGAGATATACGAGAAACAA 1122
Qy 1021 GAGCTTCACAGTTATGAGACCAACAGCAGCGGTTCATCGAACCACTGA 1071
Db 1123 GAGCTTCACAGTTATGAGACCAACAGCAGCGGTTCATCGAACCACTGA 1173

```

RESULT 5

```

AC003680/c
LOCUS
DEFINITION
AC003680
VERSION
SOURCE
KEYWORDS
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

AC003680 91854 bp DNA linear PIN 11-MAR-2002
Arabidopsis thaliana chromosome 2 BAC F17K2 genomic sequence,
complete sequence.
AC003680
AC003680.3 GI:20197048
HTG.
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 91854)
Lin X., Kaul, S., Town, C.D., Benito, M.-I., Creasy, T.H., Haas, B.J.,
Wu, D., Maiti, R., Ronning, C.M., Koo, H., Fujii, C.Y., Utterback, T.R.,
Barnstead, M.E., Bowman, C.L., White, O., Nierman, W.C. and Fraser, C.M.
Arabidopsis thaliana chromosome 2 BAC F17K2 genomic sequence
Unpublished
2 (bases 1 to 91854)
Lin X.
Direct Submission
Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
3 (bases 1 to 91854)
Town, C.D. and Kaul, S.
Direct Submission
Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA, cdtown@igr.org
On Apr 18, 2002 this sequence version replaced GI:6598396.
Address all correspondence to: atetigr.org

BAC clone F17K2 is from Arabidopsis thaliana chromosome 2 and is
near the molecular marker(s) FLS.
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of several methods: Gene
prediction programs including Genscan+ (Chris Burge).

<http://CCR-081.mit.edu/GENSCAN.html>), GeneMarkEM (Mark Borodovsky, <http://genemark.biology.gatech.edu/GeneMark/>), GlimmerA (a variant of GlimmerM, see Mihaela Pettea, <http://www.tigr.org/softlab/glimmer.htm>/glimmerM.html, and GeneSplicer (Mihaela Pettea and Steven Salzberg, contact mpettea@tigr.org), searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

FEATURES

```

source
    1..91854
        /organism="Arabidopsis thaliana"
        /mol_type="genomic DNA"
        /cultivar="Columbia"
        /db_xref="taxon:3702"
        /chromosome="2"
        /map="C1C02B07"
        /clone="F17K2"
        /complement(1..1281)
        /note="overlap with BAC clone F4L23 (AC002387:1..1281)."
        57..747
            /gene="At2g45480"
            /note="synonym: F17K2.27"
            join(<57..242,335..747)
            /gene="At2g45480"
            join(57..242,335..747)
            /gene="At2g45480"
            /note="Protein sequence is in conflict with the conceptual translation"
            /codon_start=1
            /product="hypothetical protein"
            /protein_id="AAF18607.2"
            /db_xref="GI:20197053"
            /translation="DESKPSVQKFFPEVSDKLEAKFSNNKNDIARSREKMNVM
            NGGLFHIFSPDTVLQGVETDNEPFCRGCTDQKWKRCSDVLSGGKYCD
            KIMHRGKMKHPDVTNSENAGFSPLVETAVRSVWFCKDDQDKHSVSWGTLPR
            VDEKSTSSCSTDTTITDIALRGEDDEBYLSLSPGV"
            complement(1002..2388)
            /gene="At2g45490"
            /note="synonym: F17K2.2; contains a protein kinase domain profile (PDOC00100)"
            complement(join(<1002..1112,1208..1280,1365..1580,
            1655..1750,1886..2128,2260..2388))
            /gene="At2g45490"
            complement(join(1002..1112,1208..1280,1366..1580,
            1655..1750,1886..2128,2260..2388))
            /gene="At2g45490"
            /codon_start=1
            /product="putative protein kinase"
            /protein_id="AAC06151.1"
            /db_xref="GI:2979542"
            /translation="WSKSTESDAGNTEKQWELADPEIGRLPGKKGFRVYLAREKS
            KIYALKVIFKQIEKYKHQRREMEICTSLRHPNIRLFGWFHDERIFLILEYA
            HGELYGVQNGHLEQQAAITYALSQALAYCHGKCVIHRDKPENILLDHGRLLK
            IADFGVSQSKRKTCTDLYLAPVVENRDHYADVNTWTLGILCYEFYLYGPPPE
            ARSQDKTFRIKIDILSFLPTNVSEAKNLISQILVKDPSKRSLIEKIMQHPVIVKN
            ADPKGVCSADIT"
            complement(3414..5980)
            /gene="At2g45500"
            /note="synonym: F17K2.3; predicted by genescan"
            complement(join(<3414..3616,3639..3825,3916..3965,
            4092..4215,4582..4712,4769..4864,4948..5188,5352..5374,
            5506..5585,5670..5980))
            /gene="At2g45500"

```

```

CDS
    complement(join(3414..3616,3639..3825,3916..3965,
    4092..4215,4582..4712,4769..4864,4948..5188,5352..5374,
    5506..5585,5670..5980))
    /gene="At2g45500"
    /codon_start=1
    /product="hypothetical protein"
    /protein_id="AAC06152.1"
    /db_xref="GI:2979543"
    /translation="MSFUGIIDSFSSILNEESKDPVSSSSSTSSSSNIGDVPVT
    NERIAVLYGFDLAKETIAKRAEEMGLHDDALLHYRNAQRINNEATSPSPSYIS
    SEKEVRSYREKINQWQVSRERQALCVGSENKRTVAYPSSASVSTASRYKTL
    SOKTPVARGVATPNPKDAASPRFKESGNVYDKLVEMINTTIVDRSPSVKWDVV
    AGLNGAKALLEWVILPAKRDLPFKGLREPARVTSLLGLLFGPGNGKTLAKAVAS
    EQATFNVNASLSKZKIDSIMSTRSENEASRLKSEPLIOPDGVTSNPDVAVII
    IGATNPGELDDAVLRVLVRIYVLPDSNVAKLLFKLKCQPSLSDGIDIKLVE
    TEGYVRLCKKHFISQVTDKXSGSLQALCEEAAMWPIRELGANILTIQANKVLN
    FGSVQINVEVCJSLKGLINRCVHVVTGKIST1"
    6429..8288
        /gene="At2g45510"
        /note="synonym: F17K2.4"
        join(<6429..6950,7026..7310,7386..7724,7806..8006,
        8100..8288)
        /gene="At2g45510"
        join(6429..6950,7026..7310,7386..7724,7806..8006,
        8100..8288)
        /gene="At2g45510"
        /codon_start=1
        /product="putative cytochrome P450"
        /protein_id="AAC06153.1"
        /db_xref="GI:2979544"
        /translation="MEILTSIAITVATTIFIVLCFTIYIMIRIFTKSRNDRKAPVH
        ATVFLLHSDLYDEYETIAAREKTYRFLSPQSEILTADPNVHEHLKTRFONYSK
        GHSSRENADLLGHGYFAVDGKWKQKLSGFSFSTRVLRDPSCSVFRRNASKLVGF
        VSEFALSGKFAQDLMLRCTLSIPKVGFGVELKLDGFSKEGOEWEAFDEGNVAT
        SRFDPLWKLKHPFNIGSQSKLSIATIDKPYVSLITTKRELAKQNTVREDIL
        SRFLVSEKDPENMDKYLRDIIILNFMIAKQDITAAALSWFLMLCKNPLVQEKLVQE
        IADTVSHKTTDVGTFVSEINEBALDEHYLAALSETLRLYPPVPMKCAENDDV
        LFDGHRVSGDNITYIAYANGMTYWGQDAEFPKRLKDLGFPSPKFFISFHA
        GPRICLGKDPAYQMKIVSMALLHFFRFKADENSKVYKYMULTLHVDGGLHLCAIPR
        TST"
        complement(8412..9632)
        /gene="At2g45520"
        /note="synonym: F17K2.5; supported by full length cDNA:
        Ceres:8220"
        complement(join(<8412..8678,8768..8903,8987..9120,
        9289..9356,9459..9632))
        /gene="At2g45520"
        complement(join(8553..8678,8768..8903,8987..9120,
        9289..9356,9459..9612))
        /gene="At2g45520"
        /codon_start=1
        /product="expressed protein"
        /protein_id="AAC06155.1"
        /db_xref="GI:2979546"
        /translation="WGGKGRERKNYLAHGGPARLPPLPPDRSKQDDVPSTLRIMN
        YTSFPHDSTKQVVERKEKLAKEVYVVPATSDGDDSVSEKKKRKNQMTDLAFEN
        ELAIDGRKREKREKKYWEAKKKKNGKTKETDLRENFPPKHEQFGDGVVQAPDLKA
        VVPAFKSTISASQERLQALDAYRSRKGWTFGPVPIPAVMQ"
        9408..9427
            /tpt_family="(CA)n"
            /tpt_family="(CA)n"
            complement(9578..9632)
            /tpt_family="(GA)n"
            9822..10798
                /gene="At2g45530"
                /note="synonym: F17K2.6; supported by full length cDNA:
                Ceres:35578"
                9822..10798
                    /gene="At2g45530"
                    9948..10670
                        /gene="At2g45530"
                        /codon_start=1
                        /product="expressed protein"
                        /protein_id="AAC06154.1"

```


| | | | |
|------------|------------------------------------|---|-----------------|
| DB | 74411 | ATCGCTGTCGGTTCGAATCAGCCTCAGTTATTAGCTTTTCCCGCGCGCTTCCTCGCG | 74355 |
| QY | 781 | TCGCTTTACGTCGCGCTGTTTCAACAGAGCTTCCACGATGCGTAGACACACCTTCCTTTACAA | 840 |
| DB | 74351 | TCGCTTTACGTCGCGCTGTTTCAACAGAGCTTCCACGATGCGTAGACACACCTTCCTTTACAA | 74292 |
| QY | 841 | GTTGTTTCAAGCAGCGCTTGTATCGTTCACAGCTTACGAGTTCGGTTCGAAATTTATCAAGA | 900 |
| DB | 74291 | GTTGTTTCAAGCAGCGCTTGTATCGTTCACAGCTTACGAGTTCGGTTCGAAATTTATCAAGA | 74232 |
| QY | 901 | GGCAGCTCGGTTATGGCTCCGAGCTCAAGCTCAGCGTAACAAACGGGTAGTTCATCGTCA | 960 |
| DB | 74231 | GGCAGCTCGGTTATGGCTCCGAGCTCAAGCTCAGCGTAACAAACGGGTAGTTCATCGTCA | 74172 |
| QY | 961 | ATTGCAACAACAGCAGCAGCAGCTTCCTTAGAGATATACGAGAAACAA | 1020 |
| DB | 74171 | ATTGCAACAACAGCAGCAGCAGCTTCCTTAGAGATATACGAGAAACAA | 74112 |
| QY | 1021 | GAGCTTACCACTTCATGAGCAGCAGCAGCAGCTTCGACCTGACCACTGA | 1071 |
| DB | 74111 | GAGCTTACCACTTCATGAGCAGCAGCAGCAGCTTCGACCTGACCACTGA | 74061 |
| RESULT 6 | | | |
| BT008779 | | | |
| LOCUS | | 892 bp mRNA linear | PLN 13-JUN-2003 |
| DEFINITION | | Arabidopsis thaliana At5g51910 mRNA, complete cds. | |
| ACCESSION | BT008779 | | |
| VERSION | BT008779.1 | GI:31711723 | |
| KEYWORDS | FLI CDNA. | | |
| SOURCE | Arabidopsis thaliana (thale cress) | | |
| ORGANISM | Arabidopsis thaliana | | |
| REFERENCE | | | |
| AUTHORS | | | |
| TITLE | | | |
| JOURNAL | | | |
| REFERENCE | | | |
| AUTHORS | | | |
| TITLE | | | |
| JOURNAL | | | |
| COMMENT | | | |

this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/salk) contributed equally to this work as PIs.

FEATURES
source 1..882
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="5"
/clone="U60996"
/notes="This clone is in PUN1 51
ecotype: Columbia"
1..882
/notes="unknown protein"
/codon_start=1
/product="At5g51910"
/protein_id="AAP68218.1"
/db_xref="GI:31711724"
/translation="MESNHGNAIOVDVTTWTHLSDNPKTKPGMLMKQEDGVLO
PVKTKPAKPKPTSKDRHTKVEGRIRMPAGCAARVFOITRELGHKSGETIRILLE
RAEPALIEATGTGTVPAIVSVNGTLKIPITSSVNDGGDGDGLIKRRKRNCTSD
FVDVNSCHSVTSGLAPITASNYGNILNVNTQGFVPFPMGMGTAFVTGGPDQMGQ
MMAIPTVATAPFLNVCARPVSVSYNSADAEEMETSGGTTQPLRDFSLIYDKREL
QFLGSGNSPSSCHET"
BASE COUNT 236 a 192 c 250 g 204 t
ORIGIN

Query Match 14.7%; Score 157.8; DB 8; Length 882;
Best Local Similarity 75.9%; Pred. No. 1.4e-36;
Matches 195; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
195 GACAGGACCACTTGAAGAGAGCTTCGACTAAAGACCGTCACACGAGGTTTGAAGGAAG 254
141 GACTAAACCGGCTCCGAGAGACGACTTCTTAAGACCGTCACACGAGTAGAGAGCG 200
255 AGGAGAAGGATACGATGCTCCACGCTGTCGGCTAGGATTTTCAATTAACCTCGAGA 314
201 AGGTCGAGGATCGAATGCCGCGGGTTCGCTGCTCGGGCTTTTCAATTAACCGCTGA 360
315 GTTAGGTACAAATCCGACGCGGAACGATTCGGTGTCTTCTGAGNACGCTGAGCGCGC 374
261 ACTTGGGTACAAATCCGACGCGGAACGATTCGGTGTCTTCTGAGNACGCTGAGCGCGC 320
375 GATTATAGCCGCCACGGGTACGGGAACGTTTCCGCCATCGCCATGCTGCTTAACCGAAC 434
321 GATTAATTGAACCAACCGGAACCGGAACTGTACCGGCTATTGCTGTATCGGTTAAACGAA 380
435 CTTAAAAATCCCGACGA 451
381 TTTAAAAATCCCGACGA 397

RESULT 7
AK118579 1046 bp mRNA linear PLN 06-DEC-2002
LOCUS Arabidopsis thaliana At5g51910 mRNA for unknown protein, complete cds, clone: RAFL19-81-E01.
DEFINITION AK118579
ACCESSION AK118579
VERSION AK118579.1 GI:26452181
KEYWORDS FLI_CDNA; CAP trapper.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1
Seki, M., Iida, K., Satou, M., Sakurai, T., Akiyama, K., Ishida, J., Nakajima, M., Enju, A., Kamiya, A., Narusaka, M., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.
TITLE Arabidopsis thaliana full-length cDNA
JOURNAL Published Only in Database (2002)
REFERENCE 2 (bases 1 to 1046)
AUTHORS Seki, M., Iida, K., Satou, M., Sakurai, T., Akiyama, K., Ishida, J., Nakajima, M., Enju, A., Kamiya, A., Narusaka, M., Carninci, P.,

Kawai, J., Hayashizaki, Y. and Shinozaki, K.
Direct Submission
Submitted (25-NOV-2002) Motoaki Seki, RIKEN Genomic Sciences Center; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: maseki@gsc.riken.go.jp, URL: http://pigweb.gsc.riken.go.jp, Tel: 81-45-503-9625, Fax: 81-45-503-9586)
COMMENT An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al. (1998) Plant J. 15:707-720; Seki et al. (2002) Science 296:141-145). cDNA cleaved with BamHI and XhoI was ligated to modified Lambda phage-1-E vector (Carninci et al. (2001) Genomics 77:79-90) digested with BamHI and SalI. This clone is in a modified pBluescript vector.
Please visit our web site (http://pigweb.gsc.riken.go.jp/) for further details.

FEATURES
source 1..1046
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone="RAFL19-81-E01"
/notes="common name: thale cress"
1..1046
/gene="At5g51910/MJM18.6"
47..928
/gene="At5g51910/MJM18.6"
/codon_start=1
/product="unknown protein"
/protein_id="BAC43179.1"
/db_xref="GI:26452182"
/translation="MESNHGNAIOVDVTTWTHLSDNPKTKPGMLMKQEDGVLO
PVKTKPAKPKPTSKDRHTKVEGRIRMPAGCAARVFOITRELGHKSGETIRILLE
RAEPALIEATGTGTVPAIVSVNGTLKIPITSSVNDGGDGDGLIKRRKRNCTSD
FVDVNSCHSVTSGLAPITASNYGNILNVNTQGFVPFPMGMGTAFVTGGPDQMGQ
MMAIPTVATAPFLNVCARPVSVSYNSADAEEMETSGGTTQPLRDFSLIYDKREL
QFLGSGNSPSSCHET"
BASE COUNT 277 a 208 c 287 g 274 t
ORIGIN

Query Match 14.7%; Score 157.8; DB 8; Length 1046;
Best Local Similarity 75.9%; Pred. No. 1.4e-36;
Matches 195; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
195 GACAGGACCACTTGAAGAGAGCTTCGACTAAAGACCGTCACACGAGGTTTGAAGGAAG 254
187 GACTAAACCGGCTCCGAGAGACGACTTCTTAAGACCGTCACACGAGTAGAGAGCG 246
255 AGGAGAAGGATACGATGCTTCACGCTGTCGGGCTAGGATTTTCAATTAACCTCGAGA 314
247 AGGTCGAGGATCGAATGCCGCGGGTTCGGTGTCTTCAATTAACCGCTGA 306
315 GTTAGGTACAAATCCGACGCGGAACGATTCGGTGTCTTCTGAGNACGCTGAGCGCGC 374
307 ACTTGGGTACAAATCCGACGCGGAACGATTCGGTGTCTTCTGAGNACGCTGAGCGCGC 366
375 GATTATAGCCGCCACCGGTACGGGAACGTTTCCGCCATCGCCATGCTGCTTAACGGAAC 434
367 GATTAATTGAACCAACCGGAACCGGAACGTATACCGGCTATTGCTGTATCGGTTAACGGAAC 426
435 CTTAAAAATCCCGACGA 451
427 TTTAAAAATCCCGACGA 443

RESULT 8
AY086379 1156 bp mRNA linear PLN 14-APR-2003
LOCUS Arabidopsis thaliana clone 247645 mRNA, complete sequence.
DEFINITION AY086379
ACCESSION AY086379
VERSION AY086379.1 GI:21405089
KEYWORDS FLI_CDNA.

| | | |
|------|---|-----|
| exon | /evidence=not experimental complement (251..320) /product="MADS box transcription factor-like protein" /note="CDS is reported in Acc# AB010074" gene_id:MJM18.1 /number=4 /evidence=not experimental complement (431..492) /product="MADS box transcription factor-like protein" /note="CDS is reported in Acc# AB010074" gene_id:MJM18.1 /number=3 /evidence=not experimental complement (654..738) /product="MADS box transcription factor-like protein" /note="CDS is reported in Acc# AB010074" gene_id:MJM18.1 /number=2 /evidence=not experimental complement (1636..1817) /product="MADS box transcription factor-like protein" /note="CDS is reported in Acc# AB010074" gene_id:MJM18.1 /number=1 | CDS |
| exon | /evidence=not experimental complement (3412,3636..3701,3811..3952,3943..4042,4175..4236,4360..4444,5433..5614)) /note="gene_id:MJM18.2" /codon_start=1 /evidence=not experimental /product="MADS box transcription factor-like" protein_id="BA097222.1" /db_xref="GI:8809681" /translation="MVRGTEIKKIENIVSRQVTFKRRSGLPKKAHELSVLCDQVA AIVFSQGRHLHYSOSSOMEXIIDRYGKFNFAVSRPOVERYLQELKXIEDRMVVKI DLLEVHRKLIGLQSCSVTELEIDTQIKSLRIVRSKXAEYADQIKLKEKERE LLNERKLEBVNMHSSKNTGGHRTKSSSVETDLFGLPVTFL" join(7086..7247,7585..7832,8043..8262,8344..8466) /note="gene_id:MJM18.3 unknown protein" /codon_start=1 /evidence=not experimental protein_id="BA097223.1" /db_xref="GI:8809682" /translation="MNLTNLISNLINLPKPNKMGDNGESIKWPKIKLKSNLNV SYLKNHDLFTVENCLTDSKAFVKASLGFTHQSRGPAYGEARDNHRISVNDPV LADTLWQSLNSLPTDIKIRKXAVGLNPNIRFYRSAGHPGRHIDESADLEDGRT YITLLIYLSNSTKSKSSSTNDSSAEPLVGGTETFYGSRNSIVAEVAPVEGMA LFHTHGKMLHEGRNVSKGVYFRSDVYFA" complement (join(8854..9420,9536..9724,9814..10026)) /note="gene_id:MJM18.4" /codon_start=1 /evidence=not experimental /product="peroxidase" protein_id="BA097224.1" /db_xref="GI:8809684" /translation="WATSKGLIFAMIVAVLAIKVPSEALDAHYDQSCPAARAKILLE TVRATLYDKPVARILMEFHDFCIRGCASILLDSRNOAEKDGPPNISVRSYVL TEDAKGLKACPTSCADVIAIARDVNTLGGPVSKVLEKKGDTISPRANETNL EPPIFNVSQILQSPAAGLSVKDWLTLGGHTIGFHCSEFELQNFKAFHIDISM NYAQTLLKCKPTNRGNAGVLDSTSVEDNVTYKQILSGKGVGSDQALLGDS RTKIVETFAQDQKAFREFAASVMVLGNFVKETGQVRYNTRFN" complement (join(11372..11638,11684..12145)) /note="contains similarity to cytochrome P450" gene_id:MJM18.5 /codon_start=1 /evidence=not experimental protein_id="BA097225.1" /db_xref="GI:8809684" /translation="MYRHKVPSFWKLQSMIGVGMKMIIEAGAFIDRVCGKTIASRR BEVKSQYNNNDHFTRDSDHNLTSHKLDITQYCLLDLPINDKFLRNVALLIAGRD TTSALWTFWFLSNPLVTKIQEIDMLPSCSQERPSQPMYELNKDSCMG RCRIQAREMDFRDRVSIQCRICRGQRANQKIVAVELLYQYDIKVANGQKF | CDS |
| exon | /evidence=not experimental complement (12395..13276) /note="contains similarity to DNA binding protein PCF1" gene_id:MJM18.6 /codon_start=1 /evidence=not experimental protein_id="BA097226.1" /db_xref="GI:8809685" /translation="MESNHEGNAIQVIDQVTTWTHLSDPNPKPKPMMLMKQEDGYLQ PVTKAPKRPSTSKDRTKVRGRIRMPAGCAARVOLTRELGHKSGETIRWLL RASPALEATGTGTVPALIVSVNGTLKIPTSPLVNDGGRDGDGLIKRRKRNCTSD FVDNDSCHSVTSGLAPITASNYSVNLNVTQGFVFPWPMGWTAFVTTGGPDQMG WMAIPVATAPFLAVNGARPVSSVSNASDAEAMETSGGTTQPLRDFSLIYDKREL QFJGGSGSPSSCHET" BASE COUNT 5400 a 2744 g 5153 t ORIGIN | CDS |
| exon | Query Match 14.7%; Score 157.8; DB 8; Length 16203; Best Local Similarity 75.9%; Pred. No. 2,1e-36; Matches 195; Conservative 0; Mismatches 62; Indels 0; Gaps 0; QY 195 GACAGGACCACTTGAAGAGAGCTTCGACTAAAGACCGTFCACAGAGGTTGAAGGAAG 254 DB 13136 GACTAAACCGGCTCCGAGAGAGACCGACTTCTAAAGACCGTFCACAGAGGTTGAAGGAAG 13077 QY 255 AGGAGAGAGATACGATGCTCCACAGTGTGCGGTAGGATTTTCAATTAATCACTCGAGA 314 DB 13076 AGGTCGAGAGATCGAATGCGCGGGTTCGCTGCTCGGTCTTTCAATTAATCACTCGAGA 13017 QY 315 GTTAGGTCAAAATCCACGCGGCGAAACGATTGCGTGGTTTGTGGAGAACGCTGACGCGGC 374 DB 13016 ACTTGGTCACAAATCCACGCGGAGAAACGATACGCTGATTGGAACGAGCTGAACCGGC 12957 QY 375 GATTATAGCGCCACGGGTAGCGGAACGGTTCGCGCATCGCATGTCGTTAAACGGAAC 434 DB 12956 GATAATTGAAGCAACCGGAACCGGAACGTTACCGGCTATTGCTGTATCGGTAAACGGAAC 12897 QY 435 CTTAAATAATCCCGACGA 451 DB 12896 TTTAAATAATCCCGACGA 12880 | CDS |
| exon | RESULT 10 AY103304 LOCUS Arabidopsis thaliana AT3g47620/FLP2_170 mRNA, complete cds. DEFINITION AY103304 ACCESSION AY103304.1 GI:21655288 VERSION FLI CDNA. KEYWORDS Arabidopsis thaliana (thale cress) SOURCE Arabidopsis thaliana ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis. REFERENCE AUTHORS Kim C.J., Chen, H., Cheuk, R., Meyers, M.C., Shinn, P., Banh, J., Bowser, L., Carninci, P., Chang, E., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R.W., Theologis, A. and Ecker, J.R. Arabidopsis ORF clones TITLE REFERENCE AUTHORS Kim C.J., Chen, H., Cheuk, R., Meyers, M.C., Shinn, P., Banh, J., Bowser, L., Carninci, P., Chang, E., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R.W., Theologis, A. and Ecker, J.R. Arabidopsis ORF clones TITLE REFERENCE AUTHORS | CDS |

intron 3441. .3915
/gene="FIP2.10"
/number=10
exon 3916. .4032
/gene="FIP2.10"
/number=11
intron 4033. .4173
/gene="FIP2.10"
/number=11
exon 4174. .4392
/gene="FIP2.10"
/number=12
intron 4393. .4518
/gene="FIP2.10"
/number=12
exon 4519. .4659
/gene="FIP2.10"
/number=13
intron 4660. .4736
/gene="FIP2.10"
/number=13
exon 4737. .4904
/gene="FIP2.10"
/number=14
intron 4905. .5137
/gene="FIP2.10"
/number=14
exon 5138. .5322
/gene="FIP2.10"
/number=15
intron 5323. .5490
/gene="FIP2.10"
/number=15
exon 5491. .5584
/gene="FIP2.10"
/number=16
intron 5585. .5670
/gene="FIP2.10"
/number=16
exon 5671. .5759
/gene="FIP2.10"
/number=17
intron 5760. .5841
/gene="FIP2.10"
/number=17
exon 5842. .6067
/gene="FIP2.10"
/number=18
intron 6068. .6227
/gene="FIP2.10"
/number=18
exon 6228. .6302
/gene="FIP2.10"
/number=19
intron 6303. .6412
/gene="FIP2.10"
/number=19
exon 6413. .6526
/gene="FIP2.10"
/number=20
gene 6370. .8121
/gene="FIP2.20"
gene complement(join(6970. .7100,7191. .7457,7537. .7685,
7913. .8121))
/gene="FIP2.20"
CDS complement(join(6970. .7100,7191. .7457,7537. .7685,
7913. .8121))
/gene="FIP2.20"
/note="strong similarity to CHLOROPHYLL A-B BINDING
PROTEIN 4 PRECURSOR (LHCI TYPEIII CAB-4) - Arabidopsis
thaliana, SWISSPROT:CB24_ARATH"
/codon_start=1
/product="CHLOROPHYLL A-B BINDING PROTEIN 4 PRECURSOR

homolog"

Query Match 12.0%; Score 129; DB 8; Length 101154;
Best Local Similarity 71.0%; Pred. No. 1.9e-27;
Matches 171; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 189 GTCTTCGACAGGACCAACCATTTGAAGAGAGCTTCGACTAAAGACCGTCACACGAAGTTGA 248
DB 72666 GGCAGCTAAAGACCCACCGTTTGAACGAGCGTCGCGAAGAACCGACACCGAAGATGA 72725

QY 249 AGGAAGAGGAGAGGATACGGATGCTTCCACGTGTGCGGTAGGATTTTCAATTAC 308
DB 72726 CGGAAGAGGAGGAGGATAAGGATCGCGGCTTATGTGCAGCTAGGTTTTTCACTAAC 72785

QY 309 TCGAGAGTTAGGTCACAAATCCGACGGCGAAAGGATTGGTGGTTGTCGAGAAAGCTGA 368
DB 72786 GCGAGAGCTAGGTCATAAATCCGACGCTGAGACAATAGATGGCTTCTTCAACAAGCTGA 72845

QY 369 GCGGGCGATTATAGCCGCCACCGGTACGGGAACGGTTCCCGGCATCGCCATGTCGGTTAA 428
DB 72846 ACCATCTGTAATCGCCGCCACCGAACCAGAACATCCCGGCGAATTTCACTTCTTAA 72905

QY 429 C 429
DB 72906 C 72906

RESULT 13
BX322174 451 bp DNA linear STS 10-JUN-2003
LOCUS Arabidopsis thaliana transposon insertion STS SM_3.35210, sequence
DEFINITION tagged site.
ACCESSION BX322174
VERSION BX322174.1 GI:29420305
KEYWORDS STS; STS, sequence tagged site.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE 1
AUTHORS Clarke, J.H., Bowles, B., Carter, J., Hart, D., McCullagh, B.,
Murphy, G., Langham, S., LeGrys, C., Jones, J.D.G. and Bevan, M.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 451)
AUTHORS Clarke, J.H.
TITLE Direct Submission
JOURNAL Submitted (31-MAR-2003) Clarke J.H., John Innes Centre, Colney
Lane, Norwich, NR4 7UU, UK
COMMENT AT denotes an activation tag dissociation transposon within a
single line, ET an enhancer trap dissociation transposon, GT a gene
trap dissociation transposon, MT a mis-expression enhancer trap
dissociation transposon, SM a defective suppressor mutator
transposon, _3 denotes a sequence derived from the 3' end of the
transposon, _5 denotes a sequence derived from the 5' end of the
transposon BBRC GARNET, ARIS project
On-line seed stock requests: <http://nasc.nott.ac.uk/> NASC stock
code: N121921.

FEATURES
source
1..451
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/variety="Columbia-0 NASC stock code N1092"
/db_xref="taxon:3702"
/clone="Al132965"
/note="Derived from superpool 26.24 NASC code Unknown"
1..451
/standard_name="SM_3.35210"
142 a 139 c 87 g 83 t

STS
BASE COUNT 142 a 139 c 87 g 83 t
ORIGIN

Query Match 11.6%; Score 124.6; DB 11; Length 451;
Best Local Similarity 73.1%; Pred. No. 2e-26;

Matches 160; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

2y 189 GTCTTCAGAGGACCACTTGAAGAGACTTCGACTAAGACCGCTCACAGAGTTGA 248

3b 233 GCGAGCTAAAGACCCCGTTGAACAGCGCTCGAGGAAGACCGACACAGAAAGTAGA 292

2y 249 AGGAAGAGGAGAGGATACGAGTGCCTCCAGTGTGCGGTAGGATTTTCAATTAAAC 308

3b 293 CGGAAGAGGAGGAGATTAAGATGCCGGGTATGTGCAGTAGGTTTTCACCTAAC 352

2y 309 TCAGAGTAGTTCACAATCCAGCGGGAAGAGTTCGTTGGTGTGGAGAACGCTGA 368

3b 353 GCGAGAGCTAGGTCTATAAATCCGACGGTGAGACATAGAGTGGCTTCTTCAACAAGCTGA 412

2y 369 GCGCGCGATTATAGCGCGCACGGGTACGGGAACGGTTCC 407

3b 413 ACATCTGTAAATCGCCGCCACCGGACCGGACATCCC 451

RESULT 14

LOCUS AY081706

DEFINITION Arabidopsis thaliana At1g69690/T6C23_11 mRNA, complete cds.

ACCESSION AY081706.1 GI:19547990

VERSION

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

1 (bases 1 to 978)

Cheuk, R., Chen, H., Kim, C.J., Meyers, M.C., Ban, J., Bowser, L., Carninci, P., Chang, E., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R.W., Theologis, A., and Ecker, J.R.

TITLE

Arabidopsis ORF clones

Unpublished

2 (bases 1 to 978)

Cheuk, R., Chen, H., Kim, C.J., Meyers, M.C., Ban, J., Bowser, L., Carninci, P., Chang, E., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R.W., Theologis, A., and Ecker, J.R.

Direct Submission

Submitted (28-FEB-2002) Salk Institute Genomic Analysis Laboratory (SIGNAL), Plant Biology Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

COMMENT

RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAPL cDNAs (RAPL cDNA : RIKEN Arabidopsis Full-length cDNA) : Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGSC (SSP) Consortium members carried out the sequencing and annotation of the RAPL cDNAs: Cheuk, R., Chen, H., Kim, C.J., Meyers, M.C., Shinn, P., Ban, J., Bowser, L., Chan, M.M., Chang, E., Dale, J.M., Deng, J.M., Goldsmith, A.D., Jones, T., Karlin-Neumann, G., Lam, B., Lee, J.M., Lin, J., Miranda, M., Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C., Yamada, K., Yamamura, Y., Yu, G., Davis, R.W., Theologis, A., and Ecker, J.R.

Cheuk, R. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk)

contributed equally to this work as PIs.

FEATURES

Location/Qualifiers

1..978

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/db_xref="taxon:3702"

/chromosome="1"

/clone="U14268"

/notes="This clone is in PUNI 51"

ecotype: Columbia"

CDS

1..978

/notes="unknown protein"

/codon_start=1

/product="At1g69690/T6C23_11"

/protein_id="AAL87359.1"

/db_xref="GI:19547991"

/translation="MDPDPDHNHNPFLQLDSSSSSSSLAISTTSEPNSEPKKPPKRTSTKDRHTKVEGRGRIRMPAMCAARVOLTRELGHKSDGETTETMLIQAEPAVIAATGTGTPANFTSLNISLSSSSSLAAHLRTTPSYVPHSPHSMTHLQHQHPA VRPNESHSSSSSSQLLDHNMGNLYVOSTAGSLPTQSPPATAPFWSGDNTNLMA FNINPHSGVAGDVYTPNSGGGSGVHLMNFAPIALFSGQPLASGTCGGGGGGG EHSYGVLAALNAAYRPVAETGNHNNNQNRDGHNNHNEQEDGSTSHHS"

BASE COUNT 245 a 269 c 201 g 263 t

ORIGIN

Query Match 11.6%; Score 124.6; DB 8; Length 978;

Best Local Similarity 65.0%; Pred. NO. 2.2e-26;

Matches 184; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 202 CCACCATTTGAAGAGAGCTTCGACTAAGACCGCTCACAGAGTTGAAGAGAGGAGA 261

Db 133 CTTCTCTTAAACGAACCTCTACTAAGACCGACACACCAAAAGTCGAGCGGCGCT 192

QY 262 AGATACGATGCTGCCAGCTGTGGCTAGATTTTCAATTAATCGAGATGAGT 321

Db 193 CGGATCGTATGCTGCTGCATGTGTGTGTCACGCTGCTTTCAGCTCACAGTGTGT 252

QY 322 CACAAATCCACGCGGAGAAAGATTCGTTGTTGGAGAAAGCTGAGCGGCGGATATA 381

Db 253 CACAAATCCGCGGTTGAATTTGATGGTACTTCAACAGCAGACACCGCGGTATA 312

QY 382 GCGCCACGCGGTACGGAGACGGTTCCGCCATGTCGATGCGTTAACGAACTTAAA 441

Db 313 GCGCTACAGGAGCTGACACCATTCGCGTTAACTTCTTTTAAACATCTCACTCGT 372

QY 442 ATCCCGACGCGAGCGGCTGATTTCTGATATGGTGAAATC 484

Db 373 AGCTCAAGATCTTCTCTCTGCTGCTCACTTCCTGTCACACTC 415

RESULT 15

AY056384

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

1 (bases 1 to 1440)

Cheuk, R., Chen, H., Kim, C.J., Koesena, E., Meyers, M.C., Ban, J.,

Bowser, L., Carninci, P., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y.,

Ishida, J., Jiang, P.X., Jones, T., Kamiya, A., Karlin-Neumann, G.,

Kawai, J., Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M.,

Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Pham, P.K.,

Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C.C.,

Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K.,

Davis, R.W., Theologis, A. and Ecker, J.R.

Arabidopsis cDNA clones

Unpublished

JOURNAL

REFERENCE
AUTHORS

2 (bases 1 to 1440)
 Cheuk, R., Chen, H., Kim, C.J., Koesema, E., Meyers, M.C., Banh, J.J.,
 Bowser, L., Carninci, P., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y.,
 Ishida, J., Jiang, P.X., Jones, T., Kamiya, A., Karlin-Neumann, G.,
 Kawai, J., Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M.,
 Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Pham, P.K.,
 Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C.C.,
 Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K.,
 Davis, R.W., Theologis, A. and Ecker, J.R.
 Direct Submission
 Submitted (10-SEP-2001) Salk Institute Genomic Analysis Laboratory
 (SIGAL), Plant Biology Laboratory, The Salk Institute for
 Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
 USA

COMMENT

RIKEN Genomic Sciences Center (GSC) members carried out the
 collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
 Arabidopsis Full-length cDNA') : Seki, M., Narusaka, M., Ishida, J.,
 Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
 Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGEC (SSP) Consortium members carried out the
 sequencing and annotation of the RAFL cDNAs: Cheuk, R., Chen, H.,
 Kim, C.J., Koesema, E., Meyers, M.C., Shinn, P., Banh, J., Bowser, L.,
 Dale, J.M., Goldsmith, A.D., Jiang, P.X., Jones, T., Karlin-Neumann, G.,
 Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M., Nguyen, M.,
 Onodera, C.S., Palm, C.J., Pham, P.K., Quach, H.L., Southwick, A.,
 Tang, C.C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S.,
 Davis, R.W., Theologis, A., and Ecker, J.R.

Cheuk, R. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to
 this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk)
 contributed equally to this work as PIs.

FEATURES

source

Location/Qualifiers
 1..1440
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /chromosome="1"
 /clone="RAFL09-32-L20(R14268)"
 /note="ecotype: Columbia"
 1..9

5'UTR
CDS

10..987
 /note="unknown protein"
 /codon_start=1
 /product="At1g69690/T6C23_11"
 /protein_id="AAL08240.1"
 /db_xref="GI:15912213"
 /translation="MDPPDHNHNPFLQLDSSSTSLAIITSEPNSEPKK
 PPKSTTKDRHTKVEGRIRMPACARVOLTRELGHKSDGETIEMLLQOAEPA
 VIAATGTGTPANFTSLNISRSSLSLAHLETTSSVYFHSQSMTHLOHCHQ
 VPKNESHSSSSSQLLDHNGVLYVQSTSLPTSPATAPFWSSGDNTONLWA
 FNINPHSGVAVGVYNFNSGGSGGVHLMNFAFIALFSCGPLASGTGGGGGGG
 EHSHYGVLAALNAAIRPVAETGNHNNQNRDGDHNNHNDGDSSTSHS"
 988..1440

3'UTR
BASE COUNT
ORIGIN

383 a 320 c 264 g 473 t

Query Match 11.6%; Score 124.6; DB 8; Length 1440;
 Best Local Similarity 65.0%; Pred. No. 2.3e-26;
 Matches 184; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
 y 202 CCACCAATCGAGAGAGCTTCGACTAAAGACCCTCACGAGAGGTTGAAGAGAGGAGA 261
 b 142 CTTCTCTCTAAACGAACCTCTACTAAAGACCGACACCAACCAAGTCGAAGCCGAGCCGT 201
 y 262 AGGATACGATGCTCGACAGCTGCGGCTAGGATTTTCAATTAACTCGAGAGTTAGGT 321
 b 202 CGATCCGATATGCTTGCCTATGTGCTGCAGCTGCTCTTTCAGCTCACACGTGAGCTTGGT 261
 y 322 CACAAATCCGAGCGGAGAAACGATTCGTTGTTGGAGAACGCTGAGCCGCGGATATA 381
 b 262 CACAAATCCGAGCGGTGAATACTATTGAGTGGGTACTCCAAAGCAGAACCAAGCGGTTATA 321

QY 382 GCCGCCACGGGTACGGAGACGGTTCCGCCATCGCCATGTCGGTTAAGGACCTTAAA 441
 Db 322 GCCCTACAGGACTGGAACCATTCGGGTAACTTCACTTCTTTAAACATCTCACTTCGT 381
 QY 442 ATCCCGACGACGACGACGACGCTGATTCTGATATGGGTGAAAATC 484
 Db 382 AGCTCAAGATCTTCTCTCTCTGCTGCTCATCTTCGTACACATC 424

Search completed: February 2, 2004, 00:52:51
 JOB time : 4071.41 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

M nucleic - nucleic search, using sw model

Run on: February 1, 2004, 21:36:05 ; Search time 313.739 Seconds
(without alignments)
9214.976 Million cell updates/sec

Title: US-09-938-842A-1034

Perfect score: 1071
Sequence: 1 atggcgacaattcgaagct.....cacggctcgaacacactga 1071

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: N_Geneseq_13Jun03.*
1: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/cgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/cgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/cgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/cgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/cgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDSI/cgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length DB | ID | Description |
|------------|-------|-------------|-----------|----|--------------------|
| 1 | 1071 | 100.0 | 1071 | 24 | Arabidopsis thalia |
| C 2 | 448 | 41.8 | 460 | 25 | Arabidopsis thalia |
| C 3 | 439 | 41.0 | 453 | 24 | Arabidopsis thalia |
| 4 | 157.8 | 14.7 | 1156 | 21 | Arabidopsis thalia |
| 5 | 129 | 12.0 | 1728 | 24 | Arabidopsis thalia |
| 6 | 124.6 | 11.6 | 1231 | 24 | Arabidopsis thalia |
| 7 | 124.6 | 11.6 | 1297 | 21 | Arabidopsis thalia |
| 8 | 118 | 11.0 | 1081 | 21 | A. thaliana VDBP |

| | | | | | | |
|------|-------|------|---------|----|----------|---------------------|
| 9 | 111.2 | 10.4 | 422 | 25 | ABX62217 | Arabidopsis thalia |
| 10 | 111.2 | 10.4 | 1009 | 24 | ABX65215 | Arabidopsis thalia |
| 11 | 111.2 | 10.4 | 1486 | 21 | AAC33900 | Arabidopsis thalia |
| 12 | 100 | 9.3 | 1172 | 24 | ABX65353 | Arabidopsis thalia |
| 13 | 98.6 | 9.2 | 1263 | 24 | ABX65216 | Arabidopsis thalia |
| 14 | 93.2 | 8.7 | 1379 | 21 | AAC44188 | Arabidopsis thalia |
| 15 | 89 | 8.3 | 8398 | 24 | ABX85167 | Arabidopsis thalia |
| 16 | 88.8 | 8.3 | 750 | 21 | AAC53759 | Arabidopsis thalia |
| 17 | 79.6 | 7.4 | 419 | 25 | ABX22048 | Human GDP-mannose |
| 18 | 43.6 | 4.1 | 447 | 24 | ABL93996 | Arabidopsis thalia |
| 19 | 36 | 3.4 | 2420 | 23 | ABL17625 | Drosophila melanog |
| 20 | 35.8 | 3.3 | 6161 | 23 | ABL09562 | Drosophila melanog |
| 21 | 34.6 | 3.2 | 1062 | 12 | AQ11001 | BanI restriction e |
| 22 | 34.2 | 3.2 | 8208 | 23 | ABL17624 | Drosophila melanog |
| C 23 | 34 | 3.2 | 554 | 18 | AAV74924 | Staphylococcus aur |
| C 24 | 33.6 | 3.1 | 463 | 13 | AAQ26483 | Recombinant bovine |
| C 25 | 33.6 | 3.1 | 463 | 17 | AAT35694 | Bovine (Ala47, Gly9 |
| C 26 | 33.6 | 3.1 | 3436 | 23 | ABL03290 | Drosophila melanog |
| 27 | 33.4 | 3.1 | 888 | 22 | AA194426 | Human neuroblastom |
| C 28 | 33.4 | 3.1 | 1644 | 23 | ABL16335 | Drosophila melanog |
| 29 | 33.4 | 3.1 | 4882 | 23 | ABL01868 | Drosophila melanog |
| 30 | 33.4 | 3.1 | 5675 | 23 | ABL16334 | Drosophila melanog |
| 31 | 33.2 | 3.1 | 386 | 22 | AAF65986 | Novel human polynu |
| 32 | 33.2 | 3.1 | 469 | 22 | ABA45750 | Human breast cell |
| 33 | 33.2 | 3.1 | 469 | 22 | ABA56260 | Human foetal liver |
| 34 | 33.2 | 3.1 | 469 | 22 | ABA25899 | Probe #4365 for ge |
| 35 | 33.2 | 3.1 | 469 | 22 | AAK04443 | Human brain expres |
| 36 | 33.2 | 3.1 | 469 | 22 | AAK29936 | Human bone marrow |
| 37 | 33.2 | 3.1 | 469 | 22 | AAI14529 | Probe #4462 for ge |
| 38 | 33.2 | 3.1 | 469 | 22 | AAI35902 | Probe #4588 used t |
| 39 | 33.2 | 3.1 | 469 | 22 | AAI04351 | Probe #4342 used t |
| 40 | 33.2 | 3.1 | 469 | 23 | ABS29588 | Human liver single |
| 41 | 33.2 | 3.1 | 469 | 24 | ABS04509 | Human genome-dativ |
| 42 | 33.2 | 3.1 | 1892 | 25 | ABT21247 | Aspergillus fumiga |
| C 43 | 33.2 | 3.1 | 10732 | 21 | AAA10594 | Gene encoding a su |
| 44 | 33 | 3.1 | 3011208 | 24 | ABQ69245 | Listeria innocua D |
| 45 | 32.8 | 3.1 | 4403765 | 22 | AAI99683 | Mycobacterium tube |

ALIGNMENTS

| | | |
|----------|---------------|--|
| RESULT 1 | ABZ13229 | standard; DNA; 1071 BP. |
| ID | ABZ13229 | standard; DNA; 1071 BP. |
| XX | AC | ABZ13229; |
| XX | DT | 21-JAN-2003 (first entry) |
| XX | DE | Arabidopsis thaliana stress regulated gene SEQ ID NO 1034. |
| XX | DE | Arabidopsis thaliana; plant; gene; stress; transgenic; ds. |
| XX | OS | Arabidopsis thaliana. |
| XX | XX | Arabidopsis thaliana. |
| PN | WO20021655-A2 | |
| XX | XX | 28-FEB-2002. |
| PD | XX | |
| XX | XX | 24-AUG-2001; 2001WO-US26585. |
| XX | XX | 24-AUG-2000; 2000US-227866P. |
| PR | PR | 26-JAN-2001; 2001US-264647P. |
| PR | PR | 22-JUN-2001; 2001US-300111P. |
| XX | XX | (SCRI) SCIPPS RES INST. |
| PA | PA | (SYGN) SYNGENTA PARTICIPATIONS AG. |
| XX | XX | Harper JF, Krops J, Wang X, Zhu T; |
| XX | XX | WPI; 2002-304127/34. |
| DR | XX | |

PT Identifying a stress condition to which a plant cell has been exposed
 PT and producing plants with increased tolerance to these abiotic stresses
 XX
 XX
 PS Claim 144; SEQ ID NO 1034; 577pp + Sequence Listing; English.
 CC
 CC The invention relates to identifying a stress condition to which a plant
 CC cell has been exposed, comprising:
 CC (a) contacting nucleic acid representative of expressed polynucleotides
 CC in the plant cell with an array of probes representative of the plant
 CC cell genome; and
 CC (b) detecting a profile of expressed polynucleotides in the plant cell
 CC characteristic of a stress response. The method is useful in the
 CC production of transgenic plants, cells and seeds and in producing plants
 CC with increased tolerance to abiotic stress. The present sequence is that
 CC of an Arabidopsis thaliana stress regulated gene (AB212196-AB217574) used
 CC in methods of the invention.
 CC Note: The sequence data for this patent is not represented in the printed
 CC specification but is based on sequence information supplied to Derwent by
 CC the European Patent Office.
 XX
 XX
 SQ Sequence 1071 BP; 287 A; 288 C; 256 G; 240 T; 0 other;

Query Match 100.0%; Score 1071; DB 24; Length 1071;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1071; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCGCAATTCAGAGCTTGAAGAGTTCAGGAGGCAAGATCAAACTCTTAAGAGCGTT 60
 DB 1 ATGGCGCAATTCAGAGCTTGAAGAGTTCAGGAGGCAAGATCAAACTCTTAAGAGCGTT 60
 QY 61 GATCTAACCATCATCAACGGCGTCAGAAACGTCGAAACTTCAAGACCTTCCAAAGTAAT 120
 DB 61 GATCTAACCATCATCAACGGCGTCAGAAACGTCGAAACTTCAAGACCTTCCAAAGTAAT 120
 QY 121 CCCACAGTGTCTCGAGCCCAAGGCGGACCGGTGATCGCGTGTCTTCAATGCTTTA 180
 DB 121 CCCACAGTGTCTCGAGCCCAAGGCGGACCGGTGATCGCGTGTCTTCAATGCTTTA 180
 QY 181 GCTCCACCGTCTTCGACGAGCACCACATTAAGAGAGCTTCGACTAAAGACCGTCACAG 240
 DB 181 GCTCCACCGTCTTCGACGAGCACCACATTAAGAGAGCTTCGACTAAAGACCGTCACAG 240
 QY 241 AAGTTGAAGAGAGAGGAGAGAGATACGATTCCTGCCACGTTGCGGCTAGGATTTT 300
 DB 241 AAGTTGAAGAGAGAGGAGAGAGATACGATTCCTGCCACGTTGCGGCTAGGATTTT 300
 QY 301 CAATTAACCTCGAGAGTTAGTTCACAAATCCGACGCGAAGAGTTCGCTGCGGCTAGGATTTT 360
 DB 301 CAATTAACCTCGAGAGTTAGTTCACAAATCCGACGCGAAGAGTTCGCTGCGGCTAGGATTTT 360
 QY 361 AAGCTGAGCGCGGATTAAGCGCCACGCGGTACGCGGAACGGTTCGCCCATCGCCATG 420
 DB 361 AAGCTGAGCGCGGATTAAGCGCCACGCGGTACGCGGAACGGTTCGCCCATCGCCATG 420
 QY 421 TCGGTTAACGGAACCTTAAATCCGACGAGCAGCAACGCTGATTCGATATGGTGAA 480
 DB 421 TCGGTTAACGGAACCTTAAATCCGACGAGCAGCAACGCTGATTCGATATGGTGAA 480
 QY 481 AATCTGATGAAGAGAAACGCTTAAACGAGCTTCTAAACAGTGAGTATAGACATAGCGAC 540
 DB 481 AATCTGATGAAGAGAAACGCTTAAACGAGCTTCTAAACAGTGAGTATAGACATAGCGAC 540
 QY 541 GCGGTTTACGTTCTCCCGTTTGTAGTTCCTAATGCGACGACGACGATCCAACTCGG 600
 DB 541 GCGGTTTACGTTCTCCCGTTTGTAGTTCCTAATGCGACGACGACGATCCAACTCGG 600
 QY 601 CAAGCTCTGCATCATCATCTGTGGTTCAGCAACTTCTCGCGCAAGGAATGATCCGATG 660
 DB 601 CAAGCTCTGCATCATCATCTGTGGTTCAGCAACTTCTCGCGCAAGGAATGATCCGATG 660
 QY 661 TGGGCTATTCCATCAACGCAATGATCCGAGCGTTCGAGCTTCTTCTTGTATCCACAA 720
 DB 661 TGGGCTATTCCATCAACGCAATGATCCGAGCGTTCGAGCTTCTTCTTGTATCCACAA 720

DB 661 TGGGCTATTCCATCAACGCAATGATCCGAGCGTTCGAGCTTCTTCTTGTATCCACAA 720
 QY 721 ATCGTGTGTCGTCGAATCAGCTCAGTTATAGCTTTTCCCGCGCGCTGCTTGGCGG 780
 DB 721 ATCGTGTGTCGTCGAATCAGCTCAGTTATAGCTTTTCCCGCGCGCTGCTTGGCGG 780
 QY 781 TCGTCTTACGTCGCGCTGTTTCAACAGGCTTCCACGATGCTAGACACCTCTTTACAA 840
 DB 781 TCGTCTTACGTCGCGCTGTTTCAACAGGCTTCCACGATGCTAGACACCTCTTTACAA 840
 QY 841 GTTGTTCACAGCAGCGCTTTGTATCGCTTTCAGAGCTTACGCTTTCGAATTTATCAAG 900
 DB 841 GTTGTTCACAGCAGCGCTTTGTATCGCTTTCAGAGCTTACGCTTTCGAATTTATCAAG 900
 QY 901 GCGACGTCGCTTATGCTCCGAGCTCAAGCTCAGGCGTAAACACCGGTAGTTTCATGTC 960
 DB 901 GCGACGTCGCTTATGCTCCGAGCTCAAGCTCAGGCGTAAACACCGGTAGTTTCATGTC 960
 QY 961 ATTGCAACCAACAGCAGCGCTGAGAGCTTCTCCCTAGAGATATACGAGAACAA 1020
 DB 961 ATTGCAACCAACAGCAGCGCTGAGAGCTTCTCCCTAGAGATATACGAGAACAA 1020
 QY 1021 GAGCTTCACCACTTTCATGAGCACCACCAACAGCAGCTTCATCGAACCACTGA 1071
 DB 1021 GAGCTTCACCACTTTCATGAGCACCACCAACAGCAGCTTCATCGAACCACTGA 1071

RESULT 2
 ABX61156/c
 ID ABX61156 standard; DNA; 460 BP.
 XX
 AC ABX61156;
 XX
 DT 26-FEB-2003 (first entry)
 XX
 DE Arabidopsis thaliana polynucleotide #502.
 XX
 KW Thale cress; Gene; ds; Genetic manipulation; plant; biosynthesis;
 KW Genetic modification; environmental stress; disease resistance;
 KW fungicide; insecticide; stress tolerance.
 XX
 OS Arabidopsis thaliana.
 XX
 PN US2002142319-A1.
 XX
 PD 03-OCT-2002.
 XX
 PF 07-AUG-2001; 2001US-0924035.
 XX
 PR 13-AUG-1999; 99US-148784P.
 PR 11-AUG-2000; 2000US-0638258.
 XX
 PA (GORL/) GORLACH J.
 PA (ANY/) AN Y.
 PA (HAMI/) HAMILTON C M.
 PA (PRIC/) PRICE J L.
 PA (HARG/) HARGISS T R.
 PA (YUY/) YU Y.
 PA (RAME/) RAMEKA J G.
 PA (PAGE/) PAGE A.
 PA (MATH/) MATHW A V.
 PA (LEDF/) LEDFORD B L.
 PA (WOES/) WOESSNER J P.
 PA (HAAS/) HAAS W D.
 PA (GARC/) GARCIA C A.
 XX
 PI Gorlach J, An Y, Hamilton CM, Price JL, Hargiss TR, Yu Y,
 PI Rameka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;
 PI Garcia CA;
 XX
 DR WP; 2003-102509/09.
 XX
 PT Novel Arabidopsis thaliana nucleic acid useful for constructing a

PT transgenic plant with enhanced disease resistance and enhanced traits
of interest, as probes, and in diagnosis and screening purposes -
Claim 1; Page 158; 277pp; English.
The invention relates to Arabidopsis thaliana nucleic acid sequences. The
DNA sequences and the polypeptides they encode are useful for identifying
homologous or related genes, for producing compositions that modulate the
expression or function of the polypeptides, for mapping functional
regions of the protein, in diagnosis, for studying associated
physiological pathways, for genetic manipulation of cells, preferably
plant cells, in screening assays of various plant strains to determine
the strains that are capable of withstanding a particular disease or
environmental stress, for enhancing or inhibiting production of
biosynthetic products in plants and to create genetically modified and
transgenic organisms, such as plant cells and plants. Transgenic plants
are useful for introducing or improving disease resistance and stress
tolerance in plants, screening biologically active agents, such as
fungicides and insecticides, and for identifying factors involved in
biosynthetic pathways of nutritional, commercial or medicinal value.
Sequences ABX60655-ABX61554 represent Arabidopsis thaliana
polynucleotides of the invention.
Sequence 460 BP; 113 A; 103 C; 134 G; 109 T; 1 other;
Query Match 41.8%; Score 448; DB 25; Length 460;
Best Local Similarity 99.6%; Pred No. 5e-134;
Matches 459; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
Y 600 GCAAGCTTGGCATCATCCTGCTGCTCAGCACTTCTGCGCAAGGAATGATCCGAT 659
b 460 GCAAGCTTGGCATCATCCTGCTGCTCAGCACTTCTGCGCAAGGAATGATCCGAT 401
Y 660 GGGGCTATTCCATCAAGCGCAATGATCCGACGCTGCGAGTTCTTCTGATCCACA 719
b 400 GGGGCTATTCCATCAAGCGCAATGATCCGACGCTGCGAGTTCTTCTGATCCACA 341
Y 720 AATCGTGTCTCGTGAATCAGCTTATAGCTTTTCCGCGCGCGTCTTCGCC 779
b 340 AATCGTGTCTCGTGAATCAGCTTATAGCTTTTCCGCGCGCGTCTTCGCC 282
Y 780 GTCTGTTTACGTGCGGCTGTTTCAACAGCTTCCAGATGCGTAGACCACTCTTTACA 839
b 281 GTCTGTTTACGTGCGGCTGTTTCAACAGCTTCCAGATGCGTAGACCACTCTTTACA 222
Y 840 AGTTGTTTCAACAGCGGCTTGTATCGTTTTCAGACGTTAGCGTTTCAATTTATCAAG 899
b 221 AGTTGTTTCAACAGCGGCTTGTATCGTTTTCAGACGTTAGCGTTTCAATTTATCAAG 162
Y 900 AGCGAGTGGTGTATGCTTCCAGCTCAAGCTCAGGCGTAAACCGGTAGTTTCATCGTC 959
b 161 AGCGAGTGGTGTATGCTTCCAGCTCAAGCTCAGGCGTAAACCGGTAGTTTCATCGTC 102
Y 960 AATTGCAACAACAGCGACGACGCTGAGAGCTTCTCCCTAGAGATATACGAGAACA 1019
b 101 AATTGCAACAACAGCGACGACGCTGAGAGCTTCTCCCTAGAGATATACGAGAACA 42
Y 1020 AGAGTTTCCAGTTTCAATGAGCACCACACAGCAGCGTCAAT 1060
b 41 AGAGTTTCCAGTTTCAATGAGCACCACACAGCAGCGTCAAT 1
RESULT 3
BL93850/c
D ABL93850 standard; cDNA; 453 BP.
X ABL93850;
X T 10-JUN-2002 (first entry)
X Arabidopsis thaliana nucleic acid sequence Ref:2027615 SEQ ID NO:615.
X Arabidopsis thaliana; insecticide; plant; mapping; diagnosis;

Genetic modification; Gene; ss.

Arabidopsis thaliana.

US2002023280-A1.

21-FEB-2002.

26-JAN-2001; 2001US-0770444.

27-JAN-2000; 2000US-178502P.

(GORLACH J.

(ANYI// AN Y.

(HAM// HAMILTON C M.

(PRIC// PRICE J L.

(RAIN// RAINES T M.

(YUY// YU Y.

(RAME// RAMEKA J G.

(PAGE// PAGE A.

(MATH// MATHW A V.

(LEDF// LEDFORD B L.

(WOES// WOESSNER J P.

(HAAS// HAAS W D.

(GARC// GARCIA C A.

(KRICK// KRICKER M.

(SLAT// SLATER T.

(DAVI// DAVIS K R.

(ALLE// ALLEN K.

(HOFF// HOFFMAN N.

(HURB// HURBAN P.

Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y,

Rameka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD,

Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N,

Hurban P;

WPI; 2002-267486/31.

New Arabidopsis thaliana nucleic acid, for identifying homologous

genes, producing compositions that modulate the expression or function

of its encoded protein, and mapping functional regions of a protein -

Claim 1; SEQ ID 615; 44pp; English.

The present invention describes an Arabidopsis thaliana nucleic acid (I)

comprising a sequence capable of hybridising under stringent conditions

to a sequence (S1) selected from any one of the 999 sequences given in

ABL93236 to ABL94234. (I) have insecticide and fungicide activities, and

they can be used as protein expression modulators. (I) can be used in

identifying homologous or related genes, in producing compositions that

modulate the expression or function of their encoded proteins, mapping

functional regions of the proteins, and in studying associated

physiological pathways. (I) can also be used: (1) for the genetic

manipulation of cells, particularly plant cells; (2) in screening assays

of various plant strains to determine the strains that are best capable

of withstanding a particular disease or environmental stress; (3) for

enhancing or inhibiting production of a biosynthetic product in a plant;

(4) as probes in mapping and in diagnosis, in genetic modification and

for screening purposes, to generate additional copies of the nucleic

acids, to generate ribozymes or antisense oligonucleotides, and as

single-stranded DNA probes or as triple-strand forming oligonucleotides;

and (5) for generating genetically modified transgenic organisms.

Note: The sequence data for this patent did not form part of the printed

specification, but was obtained in electronic format directly from the

USPTO web site.

Sequence 453 BP; 111 A; 100 C; 132 G; 107 T; 3 other;

Query Match 41.0%; Score 439; DB 24; Length 453;

Best Local Similarity 99.1%; Pred. No. 4e-131;

Matches 450; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 607 CTGGCATCATCACTGTGGCTCAGCAACTTCTGCGCGAAGAAATGTAATCCGATGTGGCT 666
Db 453 CTGGCATCATCACTGTGGCTCAGCAACTTCTGCGCGAAGAAATGTAATCCGATGTGGCT 394
QY 667 ATTCCATCAACGCAATGATCCGACGGTCGAGGTTTCTTCTTGATTCCACAATCGCT 726
Db 393 ATTCCATCAACGCAATGATCCGACGGTCGAGGTTTCTTCTTGATTCCACAATCGCT 334
QY 727 GGTCCGTGAATCAGCCTCAGTTATAGCTTTTCCGCGCGGCTGTCTCCGCTGTCT 786
Db 333 GGTCCGTGAATCAGCCTCAGTTATAGCTTTTCCGCGCGGCTGTCTCCGCTGTCT 275
QY 787 TACGTCGCGCTGTTTCAACAGGCTTCCAGATGGTACACCACTCTTTTACAGTGTGT 846
Db 274 TACGTCGCGCTGTTTCAACAGGCTTCCAGATGGTACACCACTCTTTTACAGTGTGT 215
QY 847 CCAACAGCGGTTTGTATCCGTTTTCAGACGTTAGCGTTTCGAATTTATCAAGACGACG 906
Db 214 CCAACAGCGGTTTGTATCCGTTTTCAGACGTTAGCGTTTCGAATTTATCAAGACGACG 155
QY 907 TCGTTATGGCTCCAGCTCAAGCTCAGGCTTAACACCGTAGTTCATCGTCAATTGCA 966
Db 154 TCGTTATGGCTCCAGCTCAAGCTCAGGCTTAACACCGTAGTTCATCGTCAATTGCA 95
QY 967 ACAACAACGACGACACGCTCAGAGACTTCTCCCTAGAGATATACGAGAACACAGCTT 1026
Db 94 ACAACAACGACGACACGCTCAGAGACTTCTCCCTAGAGATATACGAGAACACAGCTT 35
QY 1027 CACGAGTTATGAGACCAACACAGCAGCGTCAAT 1060
Db 34 CACGAGTTATGAGACCAACACAGCAGCGTCAAT 1

RESULT 4
AAC39595
ID AAC39595 standard; DNA; 1156 BP.
AC AAC39595;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 25210.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 99US-0132484.

PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134376.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144633.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.

PA (DUBE//) DUBELL A J.

PA (HEAR//) HEARD J.
PA (JIAN//) JIANG C.
PA (KEDD//) KEDDIE J.
PA (ADAM//) ADAM L.
PA (RATC//) RATCLIFF O.
PA (REUB//) REUBER J L.
PA (RIEC//) RIECHMANN J L.
PA (YUGG//) YU G.
PA (PINE//) PINEDA O.
XX
PI Pilgrim M, Creelman R, Dubell AJ, Heard J, Jiang C, Keddie J;
PI Adam L, Ratcliff O, Reuber JL, Riechmann JL, Yu G, Pineda O;
XX WPI: 2002-292022/33.
DR P-PSDB; AAU93047.
XX
XX
PT An isolated or recombinant polynucleotide used to produce a transgenic
PT plant -
XX
XX Claim 4; Page 373-376; 941pp; English.
XX
XX The invention relates to 1 of 232 isolated or recombinant polynucleotides
XX encoding an Arabidopsis thaliana transcription factor, their variants,
XX complements, fragments, or related polynucleotide with 31% to 95%
XX sequence identity, where the plant possesses an altered trait as compared
XX to a wild-type or reference plant, or the plant exhibits an altered
XX phenotype as compared to a wild-type or reference plant, or the plant
XX exhibits ectopic expression or altered expression of one or more genes
XX associated with a plant trait as compared to a wild plant. Also included
XX are a transgenic plant comprising the polynucleotides, a computer
XX readable medium having stored sequence information, and identifying a
XX homologue sequence from a database comprising a plurality of known plant
XX sequences comprising inputting sequence information selected from one of
XX 464 fully defined sequences given in the specification. The isolated or
XX recombinant polynucleotide is used for producing a plant having a
XX modified trait, the method comprising selecting a polynucleotide that
XX encodes a polypeptide or an antisense nucleic acid, inserting the
XX polynucleotide or antisense nucleic acid into an expression vector,
XX introducing the vector into a plant or a cell of a plant to overexpress
XX the polypeptide or antisense nucleic acid, thereby producing a modified
XX plant, and selecting for a modified trait (e.g. increased
XX production of agriculturally useful proteins or metabolic chemicals,
XX pest tolerance, environmental stress response (e.g. drought), microbial
XX disease resistance, herbicide resistance, seed and fruit yield, growth
XX rate, leaf and flower senescence and many other traits listed in the
XX specification). The present sequence is one of the 232 polynucleotides
XX encoding an A. thaliana transcription factor.
XX
SQ Sequence 1728 BP; 526 A; 414 C; 362 G; 426 T; 0 other;
Query Match
Best Local Similarity 12.0%; Score 129; DB 24; Length 1728;
Matches 171; Conservative 0; Mismatches 70; Indels 0; Gaps 0;
Qy 189 GTCTTCGACAGGACACCATTTGAGAGAGCTTCGACTTAAGACCGTCAACGAGGTGA 248
Db 420 GGCAGCTAAAGGACACCGCTTGAACGAGCGTCCGAGAAACGACACGACGAAAGTGA 479
Qy 249 AGGAGAGGGAGAGGATACCGATGCTTCCACCTGTGCGGTAGGATTTTCAATTAC 308
Db 480 CGGAAGGGGAGGAGAAATAGATGTCGCGGCTTATGTGACAGTAGGTTTTCAGCTAAC 539
Qy 309 TCGAGAGTAGTCAAAATCCGACGGCGAAACGATTGCTGGTGTGTGGAGACGCTGA 368
Db 540 GCGAGAGTAGTCAATAATCCGACGGTGAGACAATAGATGGCTTCTTCAACAGCTGA 599
Qy 369 GCGCGGATTATAGCCGACGAGGTACGGACCGTTCCCGCATCCCATGTCGTTAA 428
Db 600 ACCATCTGTAATCGCGCCACCGGAAACCGGAACCAATCCCGCGAATTCCTCTTTAA 659
Qy 429 C 429
Db 660 C 660

RESULT 6
ABK65217
ID ABK65217 standard; cDNA; 1231 BP.
XX
AC ABK65217;
XX
DT 02-JUL-2002 (first entry)
XX
DE Arabidopsis cDNA encoding a transcription factor #69.
XX
KW Plant; ss; gene; transcription factor; transgenic;
KW agriculture; metabolic chemical; environmental stress; drought;
KW microbial disease resistance; herbicide resistance; seed yield;
KW fruit yield; growth rate; leaf senescence; flower senescence.
XX
OS Arabidopsis thaliana.
XX
PN WO200215675-A1.
XX
PD 28-FEB-2002.
XX
PP 22-AUG-2001; 2001WO-US26189.
XX
PR 22-AUG-2000; 2000US-227439P.
PR 16-NOV-2000; 2000US-0713394.
PR 16-APR-2001; 2001US-0837944.
XX
(MEND-) MENDEL BIOTECHNOLOGY INC.
PA (PILG//) PILGRIM M.
PA (CREE//) CREELMAN R.
PA (DUBE//) DUBELL A J.
PA (HEAR//) HEARD J.
PA (JIAN//) JIANG C.
PA (KEDD//) KEDDIE J.
PA (ADAM//) ADAM L.
PA (RATC//) RATCLIFF O.
PA (REUB//) REUBER J L.
PA (RIEC//) RIECHMANN J L.
PA (YUGG//) YU G.
PA (PINE//) PINEDA O.
XX
XX Pilgrim M, Creelman R, Dubell AJ, Heard J, Jiang C, Keddie J;
XX Adam L, Ratcliff O, Reuber JL, Riechmann JL, Yu G, Pineda O;
XX WPI: 2002-292022/33.
XX P-PSDB; AAU93031.
XX
XX An isolated or recombinant polynucleotide used to produce a transgenic
XX plant -
XX
XX Claim 4; Page 315-317; 941pp; English.
XX
XX The invention relates to 1 of 232 isolated or recombinant polynucleotides
XX encoding an Arabidopsis thaliana transcription factor, their variants,
XX complements, fragments, or related polynucleotide with 31% to 95%
XX sequence identity, where the plant possesses an altered trait as compared
XX to a wild-type or reference plant, or the plant exhibits an altered
XX phenotype as compared to a wild-type or reference plant, or the plant
XX exhibits ectopic expression or altered expression of one or more genes
XX associated with a plant trait as compared to a wild plant. Also included
XX are a transgenic plant comprising the polynucleotides, a computer
XX readable medium having stored sequence information, and identifying a
XX homologue sequence from a database comprising a plurality of known plant
XX sequences comprising inputting sequence information selected from one of
XX 464 fully defined sequences given in the specification. The isolated or
XX recombinant polynucleotide is used for producing a plant having a
XX modified trait, the method comprising selecting a polynucleotide that
XX encodes a polypeptide or an antisense nucleic acid, inserting the
XX polynucleotide or antisense nucleic acid into an expression vector,
XX introducing the vector into a plant or a cell of a plant to overexpress
XX the polypeptide or antisense nucleic acid, thereby producing a modified
XX plant, and selecting for a modified trait (e.g. increased
XX production of agriculturally useful proteins or metabolic chemicals,
XX pest tolerance, environmental stress response (e.g. drought), microbial
XX disease resistance, herbicide resistance, seed and fruit yield, growth
XX rate, leaf and flower senescence and many other traits listed in the
XX specification). The present sequence is one of the 232 polynucleotides
XX encoding an A. thaliana transcription factor.

[illegible]

CC specification, but was obtained in electronic format directly from the
CC US patent office at
CC seqdata.uspto.gov/sequence.html?DocID=999909770423.
XX
SQ Sequence 422 BP; 111 A; 100 C; 118 G; 93 T; 0 other;

Query Match 10.4%; Score 111.2; DB 25; Length 422;
Best Local Similarity 71.6%; Pred. No. 3.3e-25;
Matches 146; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 209 TGRAGAGGCTTCGACTAAGACCGTCACAGAGGTTGAAGAGAGGAGAGGATAC 268
Db 57 TTAAGAAACACCGCTTAAGATCGACACAGCAAGTTGACGGAAGGAGGATTC 126

QY 269 GGATGCTCCGACGTGCGGCTAGGATTTTCAATTAACTCGAGAGTTAGGTCACAAAAT 328
Db 127 GTATGCCAATCATTTGGCAGCTCGAGTTTTCATTTGACACAGAGTTAGGTCACAAGT 186

QY 329 CCGACGGGGAACCAATTCGGTGGTTGTGGAGACGCTGAGCGGGGATATAGCGGCCA 388
Db 187 CCGATGCTCAACCAATAGAGTGGCTTCCGTCAGCTGAGCCTTCTATCATAGCGGCCA 246

QY 389 CCGGTACGGGAACGGTTCGCGCCA 412
Db 247 CTGGAATGGCACTACTCCGGCGA 270

RESULT 10
ABK65215
ID ABK65215 standard; cDNA; 1009 BP.
AC
XX
AC ABK65215;
XX
DT 02-JUL-2002 (first entry)
XX
DE Arabidopsis cDNA encoding a transcription factor #67.
XX
KW Plant; ss; gene; transcription factor; transgenic;
KW agriculture; metabolic chemical; environmental stress; drought;
KW microbial disease resistance; herbicide resistance; seed yield;
KW fruit yield; growth rate; leaf senescence; flower senescence.
XX
OS Arabidopsis thaliana.
XX
OS WO200215675-A1.
XX
XX 28-FEB-2002.
XX
XX 22-AUG-2001; 2001WO-US26189.
XX
XX 22-AUG-2000; 2000US-227439P.
XX
XX 16-NOV-2000; 2000US-0713994.
XX
XX 16-APR-2001; 2001US-0837944.
XX
XX (MEND-) MENDEL BIOTECHNOLOGY INC.
XX
XX (PILG/) PILGRIM M.
XX
XX (CREE/) CREELMAN R.
XX
XX (DUBE/) DUBELL A. J.
XX
XX (HEAR/) HEARD J.
XX
XX (JIAN/) JIANG C.
XX
XX (KEDD/) KEDDIE J.
XX
XX (ADAM/) ADAM L.
XX
XX (RATC/) RATCLIFF O.
XX
XX (REUB/) REUBER J. L.
XX
XX (RIEC/) RIECHMANN J. L.
XX
XX (YUGG/) YU G.
XX
XX (PINE/) PINEDA O.
XX
XX Pilgrim M, Creelman R, Dubell AJ, Heard J, Jiang C, Keddle J;
XX Adam L, Ratcliff O, Reuber JL, Riechmann JL, Yu G, Pineda O;
XX WPI; 2002-292022/33.
XX
XX P-PSDB; AAU93029.

XX An isolated or recombinant polynucleotide used to produce a transgenic
PT plant -
XX
XX Claim 4; Page 309-310; 941pp; English.
XX
XX The invention relates to 1 of 232 isolated or recombinant polynucleotides
CC encoding an Arabidopsis thaliana transcription factor, their variants,
CC complements, fragments, or related polynucleotide with 31% to 95%
CC sequence identity, where the plant possesses an altered trait as compared
CC to a wild-type or reference plant, or the plant exhibits an altered
CC phenotype as compared to a wild-type or reference plant, or the plant
CC exhibits ectopic expression or altered expression of one or more genes
CC associated with a plant trait as compared to a wild plant. Also included
CC are a transgenic plant comprising the polynucleotides, a computer
CC readable medium having stored sequence information, and identifying a
CC homologue sequence from a database comprising a plurality of known plant
CC sequences comprising inputting sequence information selected from one of
CC 464 fully defined sequences given in the specification. The isolated or
CC recombinant polynucleotide is used for producing a plant having a
CC modified trait, the method comprising selecting a polynucleotide that
CC encodes a polypeptide or an antisense nucleic acid, inserting the
CC polynucleotide or antisense nucleic acid into an expression vector,
CC introducing the vector into a plant or a cell of a plant to overexpress
CC the polypeptide or antisense nucleic acid, thereby producing a modified
CC plant, and selecting for a modified trait (e.g. increased
CC production of agriculturally useful proteins or metabolic chemicals,
CC pest tolerance, environmental stress response (e.g. drought), microbial
CC disease resistance, herbicide resistance, seed and fruit yield, growth
CC rate, leaf and flower senescence and many other traits listed in the
CC specification). The present sequence is one of the 232 polynucleotides
CC encoding an A. thaliana transcription factor.
XX
SQ Sequence 1009 BP; 256 A; 212 C; 267 G; 274 T; 0 other;

Query Match 10.4%; Score 111.2; DB 24; Length 1009;
Best Local Similarity 71.6%; Pred. No. 5.2e-25;
Matches 146; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 209 TGRAGAGGCTTCGACTAAGACCGTCACAGAGGTTGAAGAGAGGAGAGGATAC 268
Db 103 TTAAGAAACACCGCTTAAAGATCGACACAGCAAGTTGACGGAAGGAGGATTC 162

QY 269 GGATGCTCCGACGTGCGGCTAGGATTTTCAATTAACTCGAGAGTTAGGTCACAAAAT 328
Db 163 GTATGCCAATCATTTGGCAGCTCGAGTTTTCATTTGACACAGAGTTAGGTCACAAGT 222

QY 329 CCGACGGGGAACCAATTCGGTGGTTGTGGAGACGCTGAGCGGGGATATAGCGGCCA 388
Db 223 CCGATGCTCAACCAATAGAGTGGCTTCCGTCAGCTGAGCCTTCTATCATAGCGGCCA 282

QY 389 CCGGTACGGGAACGGTTCGCGCCA 412
Db 283 CTGGAATGGCACTACTCCGGCGA 306

RESULT 11
AAC333900
ID AAC333900 standard; DNA; 1486 BP.
XX
XX AAC333900;
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 4718.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.

XX PD 06-SEP-2000. 99US-0142055.
XX PD 06-JUL-1999; 99US-0142390.
XX PF 08-JUL-1999; 99US-0142803.
XX PF 09-JUL-1999; 99US-0142920.
XX PF 12-JUL-1999; 99US-0142977.
XX PF 13-JUL-1999; 99US-0143542.
XX PF 14-JUL-1999; 99US-0143624.
XX PF 15-JUL-1999; 99US-0144005.
XX PF 16-JUL-1999; 99US-0144085.
XX PF 16-JUL-1999; 99US-0144086.
XX PF 19-JUL-1999; 99US-0144325.
XX PF 19-JUL-1999; 99US-0144331.
XX PF 19-JUL-1999; 99US-0144332.
XX PF 19-JUL-1999; 99US-0144333.
XX PF 19-JUL-1999; 99US-0144334.
XX PF 19-JUL-1999; 99US-0144335.
XX PF 20-JUL-1999; 99US-0144352.
XX PF 20-JUL-1999; 99US-0144632.
XX PF 20-JUL-1999; 99US-0144884.
XX PF 21-JUL-1999; 99US-0144814.
XX PF 21-JUL-1999; 99US-0145086.
XX PF 21-JUL-1999; 99US-0145088.
XX PF 22-JUL-1999; 99US-0145085.
XX PF 22-JUL-1999; 99US-0145087.
XX PF 22-JUL-1999; 99US-0145089.
XX PF 22-JUL-1999; 99US-0145192.
XX PF 23-JUL-1999; 99US-0145145.
XX PF 23-JUL-1999; 99US-0145218.
XX PF 23-JUL-1999; 99US-0145224.
XX PF 26-JUL-1999; 99US-0145276.
XX PF 27-JUL-1999; 99US-0145913.
XX PF 27-JUL-1999; 99US-0145918.
XX PF 27-JUL-1999; 99US-0145919.
XX PF 28-JUL-1999; 99US-0145951.
XX PF 02-AUG-1999; 99US-0146386.
XX PF 02-AUG-1999; 99US-0146388.
XX PF 02-AUG-1999; 99US-0146389.
XX PF 03-AUG-1999; 99US-0147038.
XX PF 04-AUG-1999; 99US-0147204.
XX PF 04-AUG-1999; 99US-0147302.
XX PF 05-AUG-1999; 99US-0147192.
XX PF 05-AUG-1999; 99US-0147260.
XX PF 06-AUG-1999; 99US-0147303.
XX PF 06-AUG-1999; 99US-0147416.
XX PF 09-AUG-1999; 99US-0147493.
XX PF 09-AUG-1999; 99US-0147935.
XX PF 10-AUG-1999; 99US-0148171.
XX PF 11-AUG-1999; 99US-0148319.
XX PF 12-AUG-1999; 99US-0148341.
XX PF 13-AUG-1999; 99US-0148565.
XX PF 13-AUG-1999; 99US-0148684.
XX PF 16-AUG-1999; 99US-0149368.
XX PF 17-AUG-1999; 99US-0149175.
XX PF 18-AUG-1999; 99US-0149426.
XX PF 20-AUG-1999; 99US-0149722.
XX PF 20-AUG-1999; 99US-0149723.
XX PF 20-AUG-1999; 99US-0149929.
XX PF 23-AUG-1999; 99US-0149902.
XX PF 23-AUG-1999; 99US-0149930.
XX PF 25-AUG-1999; 99US-0150566.
XX PF 26-AUG-1999; 99US-0150884.
XX PF 27-AUG-1999; 99US-0151065.
XX PF 27-AUG-1999; 99US-0151066.
XX PF 27-AUG-1999; 99US-0151080.
XX PF 30-AUG-1999; 99US-0151303.
XX PF 31-AUG-1999; 99US-0151438.
XX PF 01-SEP-1999; 99US-0151930.
XX PF 07-SEP-1999; 99US-0152363.
XX PF 10-SEP-1999; 99US-0153070.
XX PF 13-SEP-1999; 99US-0153758.
XX PF 15-SEP-1999; 99US-0154018.
XX PF 16-SEP-1999; 99US-0154039.
XX PF 20-SEP-1999; 99US-0154779.

22-SEP-1999; 99US-01551139.
23-SEP-1999; 99US-0155486.
24-SEP-1999; 99US-0155659.
28-SEP-1999; 99US-0156458.
29-SEP-1999; 99US-0156596.
04-OCT-1999; 99US-0157117.
05-OCT-1999; 99US-0157753.
06-OCT-1999; 99US-0157865.
07-OCT-1999; 99US-0158029.
08-OCT-1999; 99US-0158232.
12-OCT-1999; 99US-0158369.
13-OCT-1999; 99US-0159293.
13-OCT-1999; 99US-0159295.
14-OCT-1999; 99US-0159329.
14-OCT-1999; 99US-0159330.
14-OCT-1999; 99US-0159331.
14-OCT-1999; 99US-0159637.
14-OCT-1999; 99US-0159638.
18-OCT-1999; 99US-0159584.
21-OCT-1999; 99US-0160741.
21-OCT-1999; 99US-0160767.
21-OCT-1999; 99US-0160768.
21-OCT-1999; 99US-0160770.
21-OCT-1999; 99US-0160814.
21-OCT-1999; 99US-0160815.
22-OCT-1999; 99US-0160980.
22-OCT-1999; 99US-0160981.
22-OCT-1999; 99US-0160989.
25-OCT-1999; 99US-0161404.
25-OCT-1999; 99US-0161405.
25-OCT-1999; 99US-0161406.
26-OCT-1999; 99US-0161339.
26-OCT-1999; 99US-0161360.
26-OCT-1999; 99US-0161361.
28-OCT-1999; 99US-0161920.
28-OCT-1999; 99US-0161992.
28-OCT-1999; 99US-0161993.
29-OCT-1999; 99US-0162142.
29-OCT-1999; 99US-0162142.

Query Match 10.4%; Score 111.2; DB 21; Length 1486;
Best Local Similarity 71.6%; Pred. No. 6.4e-25;
Matches 146; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 209 TGAAGAGAGCTTCACAAACGCTCACAGAGGTTGAAGGAGGAGGAGGATC 268
Db 150 TTAGAAACCAACCGTCTAAGATCGACACAAAGTTGACGGAGGAGGAGGATC 209
QY 269 GGATGCCCTGCCACGTGTGCGCTAGGATTTTCAATTAACTCGAGAGTTAGGTCACAAAAT 328
Db 210 GTATGCCAATCATTTGGCAGCTCGAGTTTTCATTTGACCAAGAGTTAGGTCACAAGT 269
QY 329 CGGACGGGAACAGATTCGGTGTGTTGGAGACGCTGAGCGGCGGATTAAGCGGCA 388
Db 270 CCGATGGTCAACACCATAGATGGGCTTTCGTCGAAGTGAGCCTTCTATCATAGCGGCA 329
QY 389 CGGATACGGGAACGTTCCCGCCA 412
Db 330 CTGGACTGGCACTACCTCCGCGA 353

RESULT 12
ABK65353
ID ABK65353 standard; cDNA; 1172 BP.
XX AC ABK65353;
XX AC ABK65353;
DT 02-JUL-2002 (first entry)
XX Arabidopsis cDNA encoding a transcription factor #205.
XX Plant; ss: gene; transcription factor; transgenic;
KW agriculture; metabolic chemical; environmental stress; drought;

KW microbial disease resistance; herbicide resistance; seed yield;
KW fruit yield; growth rate; leaf senescence; flower senescence.
XX Arabidopsis thaliana.
OS WO200215675-A1.
FN 28-FEB-2002.
XX 22-AUG-2001; 2001WO-US26189.
XX 22-AUG-2000; 2000US-227439P.
PR 16-NOV-2000; 2000US-0713994.
PR 16-APR-2001; 2001US-0837944.
XX (MEND-) MENDEL BIOTECHNOLOGY INC.
PA (PILG/) PILGRIM M.
PA (CREE/) CREELMAN R.
PA (DUBE/) DUBELL A J.
PA (HEAR/) HEARD J.
PA (JIAN/) JIANG C.
PA (KEDD/) KEDDIE J.
PA (ADAM/) ADAM L.
PA (RATC/) RATCLIFF O.
PA (REUB/) REUBER J L.
PA (RIEC/) RIECHMANN J L.
PA (YUGG/) YU G.
PA (PINE/) PINEDA O.
XX Pilgrim M, Creelman R, Dubell AJ, Heard J, Jiang C, Keddie J;
PI Adam L, Ratcliff O, Reuber JL, Riechmann JL, Yu G, Pineda O;
DR WPI; 2002-292022/33.
DR P-PSDB; AAU93167.
XX An isolated or recombinant polynucleotide used to produce a transgenic
PT plant -
XX Claim 4; Page 845-846; 941pp; English.
XX The invention relates to 1 of 232 isolated or recombinant polynucleotides
CC encoding an Arabidopsis thaliana transcription factor, their variants,
CC complements, fragments, or related polynucleotide with 31% to 95%
CC sequence identity, where the plant possesses an altered trait as compared
CC to a wild-type or reference plant, or the plant exhibits an altered
CC phenotype as compared to a wild-type or reference plant, or the plant
CC exhibits ectopic expression or altered expression of one or more genes
CC associated with a plant trait as compared to a wild plant. Also included
CC are a transgenic plant comprising the polynucleotides, a computer
CC readable medium having stored sequence information, and identifying a
CC homologue sequence from a database comprising a plurality of known plant
CC sequences comprising inputting sequence information selected from one of
CC 464 fully defined sequences given in the specification. The isolated or
CC recombinant polynucleotide is used for producing a plant having a
CC modified trait, the method comprising selecting a polynucleotide that
CC encodes a polypeptide or an antisense nucleic acid, inserting the
CC polynucleotide or antisense nucleic acid into an expression vector,
CC introducing the vector into a plant or a cell of a plant to overexpress
CC the polypeptide or antisense nucleic acid, thereby producing a modified
CC plant, and selecting for a modified trait (e.g. increased
CC production of agriculturally useful proteins or metabolic chemicals,
CC pest tolerance, environmental stress response (e.g. drought), microbial
CC disease resistance, herbicide resistance, seed and fruit yield, growth
CC rate, leaf and flower senescence and many other traits listed in the
CC specification). The present sequence is one of the 232 polynucleotides
CC encoding an A. thaliana transcription factor.
XX Sequence 1172 BP; 331 A; 217 C; 251 G; 373 T; 0 other;
SQ

Query Match 9.3%; Score 100; DB 24; Length 1172;
Best Local Similarity 68.1%; Pred. No. 2.3e-21;
Matches 139; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

XX PR 25-FEB-2000; 2000EP-0301439. 99US-0142803.
XX PF 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149436.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149923.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.

R 24-SEP-1999; 99US-0155659.
R 28-SEP-1999; 99US-0156458.
R 29-SEP-1999; 99US-0156596.
R 04-OCT-1999; 99US-0157117.
R 05-OCT-1999; 99US-0157753.
R 06-OCT-1999; 99US-0157865.
R 07-OCT-1999; 99US-0158029.
R 08-OCT-1999; 99US-0158232.
R 12-OCT-1999; 99US-0158369.
R 13-OCT-1999; 99US-0159293.
R 13-OCT-1999; 99US-0159294.
R 13-OCT-1999; 99US-0159295.
R 14-OCT-1999; 99US-0159329.
R 14-OCT-1999; 99US-0159330.
R 14-OCT-1999; 99US-0159331.
R 14-OCT-1999; 99US-0159637.
R 14-OCT-1999; 99US-0159638.
R 18-OCT-1999; 99US-0159584.
R 21-OCT-1999; 99US-0160741.
R 21-OCT-1999; 99US-0160767.
R 21-OCT-1999; 99US-0160768.
R 21-OCT-1999; 99US-0160770.
R 21-OCT-1999; 99US-0160814.
R 21-OCT-1999; 99US-0160815.
R 22-OCT-1999; 99US-0160980.
R 22-OCT-1999; 99US-0160981.
R 22-OCT-1999; 99US-0160989.
R 25-OCT-1999; 99US-0161404.
R 25-OCT-1999; 99US-0161405.
R 25-OCT-1999; 99US-0161406.
R 26-OCT-1999; 99US-0161359.
R 26-OCT-1999; 99US-0161360.
R 26-OCT-1999; 99US-0161361.
R 28-OCT-1999; 99US-0161920.
R 28-OCT-1999; 99US-0161992.
R 28-OCT-1999; 99US-0161993.
R 29-OCT-1999; 99US-0162142.

Query Match 8.7%; Score 93.2; DB 21; Length 1379;
Best Local Similarity 68.8%; Pred. No. 48-19; 58; Indels 0; Gaps 0;
Matches 128; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
Y 209 TGAAGAGAGCTTCGACTAAGACCCGTACACGAGAGTTGAAGAGAGAGGAGAGGATAC 268
b 195 TGAAGAAACACCGCGGAAGATCGACATAGCAAGTGTGATGAGAGAGGAGAGATCC 254
Y 269 GATGCTCCACGCTGTCGGCTAGATTTTCAATTAATTAATTAATTAATTAATTAATTAAT 328
b 255 GTATGCCGATTATATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 314
Y 329 CCGACGGCGAAGATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 388
b 315 CAGATGCCAACAATGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 374
Y 389 CCGGTA 394
b 375 CAGGGA 380

RESULT 15:
ABN85767
ID ABN85767 standard; cDNA; 83698 BP.
XX AC ABN85767;
XX 21-OCT-2002 (first entry)
XX Arabidopsis yellow stripe-like 4 encoding cDNA SEQ ID NO 9.
XX Maize; transgenic; plant; yellow stripe-like; ysl; Arabidopsis;
XX iron uptake, bioremediation; yellow stripe 1; ysl; gene; ss.
XX Arabidopsis sp.

XX WO200240688-A2.
XX 23-MAY-2002.
XX 16-NOV-2001; 2001WO-US43101.
XX 16-NOV-2000; 2000US-249222P.
XX (UYUA) UNIV YALE.
XX Walker EL, Dellaporta S;
XX WPI; 2002-490144/52.
XX P-PSDB; AB83920.
XX New yellow stripe and yellow stripe-like genes, useful for altering
XX the distribution of iron within the plant body so that edible parts of
XX crop plants have more iron, or for producing plants useful in enhancing
XX iron uptake from soil -
XX Claim 1; Page 114-159; 187pp; English.
XX The invention relates to an isolated nucleic acid molecule (I), maize
XX yellow stripe 1 (ysl) or yellow stripe-like (ysl) from Arabidopsis
XX (ABN85763-ABN85771). (I) is useful for generating transgenic plants which
XX can be used for enhancing iron uptake from soil and for bioremediation of
XX metal or heavy metal contaminated soil. (I) may also be used to alter the
XX distribution of iron within the plant body so that edible parts of crop
XX plants have more iron. Transgenic plants may also be used in conventional
XX plant breeding schemes to produce progeny which also contain the gene of
XX interest. The present sequence is that of the Arabidopsis ysl encoding
XX cDNA of the invention.
XX SQ Sequence 83698 BP; 26295 A; 15150 C; 15104 G; 27149 T; 0 other;

Query Match 8.3%; Score 89; DB 24; Length 83698;
Best Local Similarity 65.2%; Pred. No. 8.3e-17;
Matches 131; Conservative 0; Mismatches 70; Indels 0; Gaps 0;
Y 210 GAAGAGAGCTTCGACTAAGACCCGTACACGAGGTTGAAGAGAGGAGGAGGATACG 269
b 18681 GAAGAAGAACCAACAAAGATCGTCCTTAAGTTGAAGAGAGGTCGTAGAGTTAG 18740
Y 270 GATGCTCCACCGCTGTCGGCTAGGATTTTCAATTAATTAATTAATTAATTAATTAAT 329
b 18741 GTTACCTCCACTCTGTCAGCAAGGATTTTCAATTAATTAATTAATTAATTAATTAAT 18800
Y 330 CGACGGCGAAGCAATTCGGTGGTTGAGAGACCGTGAAGCGCGCGGATTAAGCGCCAC 389
b 18801 AGATGGTGAGACTCTTGAATGGTTGCTTCAACATGCTGAGCCATCCTCTGCTAC 18860
Y 390 GGGTACGGGAACGGTTCCCGC 410
b 18861 TGTAAATGTTATCAAAACCCAC 18881

Search completed: February 1, 2004, 21:58:51
Job time : 322.739 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2004, 21:45:16 / Search time 70.5426 Seconds
(without alignments)
6701.220 Million cell updates/sec

Title: US-09-938-842A-1034

Perfect score: 1071

Sequence: 1 atggcgacaattcgaagct.....cagcgatcgcgaaccactga 1071

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

- 1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCRTUS.COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|---------|----|----------------------|
| 1 | 44 | 4.1 | 7218 | 1 | US-08-232-463-14 |
| 2 | 41 | 3.8 | 7218 | 1 | US-08-232-463-14 |
| 3 | 33 | 3.1 | 2127 | 4 | US-09-252-991A-8192 |
| 4 | 33 | 3.1 | 2874 | 4 | US-09-252-991A-8112 |
| 5 | 32.8 | 3.1 | 4403765 | 3 | US-09-103-840A-2 |
| 6 | 32.2 | 3.0 | 2406 | 4 | US-08-632-098-5 |
| 7 | 32.2 | 3.0 | 2439 | 4 | US-08-632-098-5 |
| 8 | 31.4 | 2.9 | 4411529 | 3 | US-09-103-840A-1 |
| 9 | 30.6 | 2.9 | 412 | 3 | US-08-961-083-111 |
| 10 | 30.6 | 2.9 | 412 | 4 | US-09-536-784-111 |
| 11 | 30.6 | 2.9 | 1288 | 4 | US-09-620-312D-546 |
| 12 | 30.6 | 2.9 | 6693 | 4 | US-08-961-527-195 |
| 13 | 30.4 | 2.8 | 2172 | 1 | US-07-982-712-1 |
| 14 | 30.4 | 2.8 | 7766 | 4 | US-08-135-619-3 |
| 15 | 30.4 | 2.8 | 580073 | 4 | US-08-585-828D-1 |
| 16 | 30.2 | 2.8 | 801 | 3 | US-08-998-416-436 |
| 17 | 30.2 | 2.8 | 1553 | 3 | US-09-217-490-1 |
| 18 | 30.2 | 2.8 | 2396 | 4 | US-09-221-017B-74 |
| 19 | 30.2 | 2.8 | 23673 | 4 | US-09-773-816-1 |
| 20 | 30 | 2.8 | 536 | 3 | US-08-714-918-57 |
| 21 | 30 | 2.8 | 536 | 3 | US-09-265-315-57 |
| 22 | 30 | 2.8 | 536 | 3 | US-09-265-315-57 |
| 23 | 30 | 2.8 | 536 | 3 | US-09-266-417-57 |
| 24 | 30 | 2.8 | 1664976 | 4 | US-08-916-421B-1 |
| 25 | 29.8 | 2.8 | 582 | 4 | US-09-252-991A-12349 |
| 26 | 29.8 | 2.8 | 618 | 4 | US-09-252-991A-12199 |
| 27 | 29.8 | 2.8 | 1026 | 4 | US-09-252-991A-12295 |

Sequence 5, Appli
Sequence 348, Appl
Sequence 2, Appli
Sequence 1, Appli
Sequence 79, Appli
Sequence 2, Appli
Sequence 1, Appli
Sequence 3, Appli
Sequence 6, Appli
Sequence 3, Appli
Sequence 1105, Ap
Sequence 1, Appli
Sequence 11, Appli
Sequence 5277, Ap
Sequence 17, Appl
Sequence 2, Appli
Sequence 2, Appli

28 29.8 2.8 2628 4 US-09-294-531B-5
29 29.8 2.8 4062 4 US-09-520-512B-348
30 29.8 2.8 8878 1 US-08-759-444-2
31 29.8 2.8 9880 3 US-08-680-897-1
32 29.8 2.8 38155 4 US-09-453-702B-79
33 29.8 2.8 4403765 3 US-09-103-840A-2
34 29.8 2.8 4411529 3 US-09-103-840A-1
35 29.6 2.8 1302 2 US-08-529-600D-3
36 29.6 2.8 1302 2 US-08-973-275-6
37 29.6 2.8 1302 2 US-09-122-632-3
38 29.6 2.8 4065 4 US-09-016-434-1105
39 29.6 2.8 4739 3 US-08-585-871-1
40 29.6 2.8 3960 4 US-09-453-702B-11
41 29.4 2.7 278 4 US-09-313-294A-5277
42 29.4 2.7 289 3 US-09-007-005-17
43 29.4 2.7 289 3 US-09-244-796-17
44 29.4 2.7 2277 1 US-08-676-967-2
45 29.4 2.7 2277 1 US-08-676-974-2

ALIGNMENTS

RESULT 1

US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
PRIOR APPLICATION DATA:
CLASSIFICATION: 435
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMD
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZsept-Fls
US-08-232-463-14

Query Match 4.1%; Score 44; DB 1; Length 7218;

```

Best Local Similarity 2.3%; Pred. No. 0.0011;
Matches 8; Conservative 203; Mismatches 143; Indels 0; Gaps 0;

QY 542 CCGTTTCAGTCTCCTCGGTTAGCTTCAATGCCACGACGACGATCCAACTCCGC 601
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1093 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1152
QY 602 AAGCTCTGGCATCATCCTGCTGCTCAGCACTCTCGCGAAGATGATCCGATGT 661
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1153 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1212
QY 662 GGCTATTCATCAACGAAATGATTCGACGCTCGAGCTTCTTCTGATTCACAAA 721
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1213 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1272
QY 722 TCGCTGCTCGGTCGATCAGCTCAGTATTAGCTTTCCCGCCGCTGCTCCGCGT 781
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1273 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1332
QY 782 CGTCTTACGTCGCGCTGTTCAACAGGCTTCACGATGCTAGACACCTCTTACAAG 841
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1333 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1392
QY 842 TTGTTCCAGCAGCGCTTGTATCGTTTCAGCTTTCAGCTTTCGATTTAT 895
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1393 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1446

```

RESULT 2

```

US-08-232-463-14/C
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500.
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 INMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

```

```

IMMEDIATE SOURCE:
CLONE: ptzgt-Fls
US-08-232-463-14

```

```

Query Match 3.8%; Score 41; DB 1; Length 7218;
Best Local Similarity 2.2%; Pred. No. 0.01;
Matches 8; Conservative 205; Mismatches 150; Indels 0; Gaps 0;

QY 195 GACGAGACCACTTGAAGAGAGCTTCGACTAAAGACCTCACAGAGTTGAAGAAAG 254
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1414 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1355
QY 255 AGGAGAGATACGATCGCTCCACAGTGTGGGTAGGATTTTCAATTAACTCGAGA 314
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1354 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1295
QY 315 GTTAGTGTCAAAATCCGACGCGGAAACAGTTCGGTGTGTTGGAGAACGCTGACCGC 374
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1294 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1235
QY 375 GATTATAGCGGCACGGGTACGGGAACGGTTCGCCCATCGCCATGTCGGTTAAACGAA 434
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1234 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1175
QY 435 CTAAAAATCCGACGACGACGACGCTGATTCGATATGATGGTGAATCTGATGAAGA 494
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1174 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1115
QY 495 GAAACGTAACGACCTTCAACAGTGAAGTATATAGACGACGCGGTTTCAGCTTC 554
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1114 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1055
QY 555 CTC 557
Db : : : : :
1054 CTC 1052

```

RESULT 3

```

US-09-252-991A-8192
Sequence 8192, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 8192
LENGTH: 2127
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
FEATURE:
NAME/KEY: unsure
LOCATION: (291)
OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.
US-09-252-991A-8192

```

```

Query Match 3.1%; Score 33; DB 4; Length 2127;
Best Local Similarity 49.2%; Pred. No. 2.1;
Matches 87; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 111 CCAAGTAAATCCACAGTGAAGTCTCGAGCCCAAGCGGAGCGGATGCGGTTC 170
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
70 CCAAGTGAACAGCTGGAGCGCTATAGACACGACCGCCGAGAGCGCCCTGAGCACC 129
QY 171 AATGCTTTAGTCCACCGTCTTCGACAGGACCACTTGAAGAGAGCTTCGATAAGA 230

```

Db 130 AATACCGGACCCGATCGCGGATACCAAGAACACCCCTGAAGCCGCGGAGCGGACCG 189
Qy 231 CCCTCACACGAAGTTGAAGAAAGAGGAGGATACCGATCCCTGCCACGCTGTGC 287
Db 190 TCGTGTCTGAAGACTTCATCATGCGGAGAGATCACCCACTTCGACCAACGAGCGC 246

RESULT 4

US-09-252-991A-8112/c
; Sequence 8112, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8112
; LENGTH: 2874
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (2266)
; OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.
US-09-252-991A-8112

Query Match 3.1%; Score 33; DB 4; Length 2874;

Best Local Similarity 49.2%; Pred. No. 2.5;
Matches 87; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

Qy 111 CCAAGTAAATCCACAGTGTCTCGAGCCCAAGGCGGAGCGGTGATCGCTGCTTTC 170
Db 2487 CCAAGCTGGAACAGCTGGAGGCTATAGCAGACGACGCCACCGGAGGCGCTGAGCACC 2428
Qy 171 AATGCTTTAGTCCACCGTCTTCGACAGACACCACTTGAAGAGAGCTTCGACTAAAGA 230
Db 2427 AATACCGGACCCGATCGCGATACCAAGAACACCCCTGAAGCCGCGGAGCGGACCG 2368
Qy 231 CCCTCACACGAAGTTGAAGAAAGAGGAGGATACCGATCCCTGCCACGCTGTGC 287
Db 2367 TCGTGTCTGAAGACTTCATCATGCGGAGAGATCACCCACTTCGACCAACGAGCGC 2311

RESULT 5

US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence

; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 3.1%; Score 32.8; DB 3; Length 4403765;
Best Local Similarity 59.8%; Pred. No. 68;
Matches 55; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 133 CTCGAGCCCAAGCGGAGCGCGTGTATCGCTGCTTTTCAATGCTTTAGCTCCACCGTCT 192
Db 3941480 CTGCGGACAAAGCGCGCGCGGTGCTGCGGGCGCGCGGATTAACCCACCGGCA 3941539
Qy 193 TCGACAGGACCACTTGAAGAGAGCTTCGAC 224
Db 3941540 TCGCGGACCGCGGTGTCGCGGCGCACCGGC 3941571

RESULT 6

US-09-632-098-5
; Sequence 5, Application US/09632098
; Patent No. 6420154
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Balindur, Nand
; APPLICANT: Bishop, Paul D.
; TITLE OF INVENTION: MAMMALIAN ADHESION PROTEASE PEPTIDES
; FILE REFERENCE: 99-39
; CURRENT APPLICATION NUMBER: US/09/632,098
; CURRENT FILING DATE: 2000-08-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 2406
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(2406)
; OTHER INFORMATION: n = A, T, C or G
US-09-632-098-5

Query Match 3.0%; Score 32.2; DB 4; Length 2406;

Best Local Similarity 25.2%; Pred. No. 4.1;
Matches 82; Conservative 47; Mismatches 197; Indels 0; Gaps 0;

Qy 637 CTGCGGCAAGAAATGATCCGATGTGGCTATTCCATCAAGCAATGATTCGAGCGTC 696
Db 1016 CNGCNGCNAATGGCNCAYGAATHGNCAYNSNTNGNNTNWSNCAYGAYCNCNGAYG 1075
Qy 697 GGAGCTTTCTTCTTGATTCCACAAATCGCTGTGCTCCGATCAGCTCAGTTATTAGCT 756
Db 1076 GNTGTYGTNGARGCNGCNGCNGARWSNGGNGTGYTNATGGCNGCNGCNGCNGC 1135
Qy 757 TTTCCCGCCCGCTGCTTCGCCGTGCTTACGTGCTGCTGCTTCAACAGGCTCCAGC 816
Db 1136 AYCNTTYCCNMGNGTNTTYWSNGCNTGYWSNMGNCARYTNMGNCNTTYTYMGNA 1195
Qy 817 ATGGCTAGACCACTCTTTTACAAGTTGTTCCAAAGCAGCGGCTTTGTATCCGTTTCAGAC 876
Db 1196 ARGNGGNGGNGCNTGYTNWSNAAAGCNCNGAYCNGGNTNCCNGCNGCNGNY 1255
Qy 877 GTTAGCGTTTGAATTTATCAAGAGCGAGCTCGGTTATGGCTCCGAGCTCAAGCTCAGC 936
Db 1256 TMTGYGNAAYGNTTYGTNGARGCNGGNGARGARTGYGAYTYGNGCNGCNGCARGART 1315
Qy 937 GTAACACCGGTAGTTTCATCGTCAAT 962
Db 1316 GYMNGAYYNTGYTYTTCNCAY 1341

RESULT 7

US-09-632-098-6
; Sequence 6, Application US/09632098

[illegible]

GENERAL INFORMATION:

APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/536,784
FILING DATE: 30-Oct-1997
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997

ATTORNEY/AGENT INFORMATION:

NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 111:

SEQUENCE CHARACTERISTICS:

LENGTH: 412 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 111:

JS-09-536-784-111

Query Match

Best Local Similarity 2.9%; Score 30.6; DB 4; Length 412;

Matches 63; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

2Y 798 TGTTCACAGGCTTCACAGTGTGACACCTCTTTACAGTGTGTTCCAGGCGG 857

Db 177 TGTCCAACTGCTCTTTATCATGTGCGACTGCGTCCACAGACCTTTTCAAAGGGGG 236

2Y 858 CTTTGTATCGTTTCAGAGCTTAGCGTTTCGAATTTATCAAGAGCGAGTCGGTTAT 914

Db 237 CTTTGTCTAGTACCGCTGTTATTCCTCGCGGTGTGTCAGAGTGCCGTGCTTTAT 293

RESULT 11

JS-09-620-312D-546/c

Sequence 546, Application US/09620312D

Patent No. 6569562

GENERAL INFORMATION:

APPLICANT: Tang, Y. Tom

APPLICANT: Liu, Chenghua

APPLICANT: Asundi, Vinod

APPLICANT: Zhang, Jie

APPLICANT: Ren, Feiyan

APPLICANT: Chen, Rui-hong

APPLICANT: Zhao, Qing A.

APPLICANT: Wehrman, Tom

APPLICANT: Xue, Aidong J.

APPLICANT: Yang, Yonghong

APPLICANT: Wang, Jian-Rui

APPLICANT: Zhou, Ping

APPLICANT: Ma, Yunding

APPLICANT: Wang, Dunrui

APPLICANT: Wang, Zhiwei

APPLICANT: John Tillinghast

APPLICANT: Drmanac, Radoje T.

TITLE OF INVENTION: No. 6569662e1 Nucleic Acids and

FILE REFERENCE: Polypeptides

CURRENT APPLICATION NUMBER: US/09/620,312D

PRIOR FILING DATE: 2000-07-19

PRIOR FILING DATE: 2000-04-25

PRIOR FILING DATE: 2000-01-21

NUMBER OF SEQ ID NOS: 1105

SOFTWARE: pt FL_genes Version 1.0

SEQ ID NO 546

LENGTH: 1288

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (95)...(928)

US-09-620-312D-546

Query Match 2.9%; Score 30.6; DB 4; Length 1288;

Best Local Similarity 55.0%; Pred. No. 9.9;

Matches 60; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 7 ACATTTCAGAGCTTGAAGAGTTGCAGGCAAGATCAAACTTAAGAGCGGTTGATCTA 66

Db 439 ATAGTCGGACCTCCACCTTTTGGCTCCAGATGTTCTTCAGAGTCTCTGTCATCAA 380

Qy 67 ACCATCATCAACGGCGTGCAGAAACGTCGAAACTTCAAGACCTTTTCAAG 115

Db 379 ATCATTATCAAGGCATCTGGATGGTCAAAATTTGACTGTCTTTTCAAG 331

RESULT 12

US-08-961-527-195/c

Sequence 195, Application US/08961527

Patent No. 6420135

GENERAL INFORMATION:

APPLICANT: Charles Kunsch

TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

NUMBER OF SEQUENCES: 391

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961,527

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 195:

SEQUENCE CHARACTERISTICS:

LENGTH: 6693 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear
US-08-961-527-195

Query Match 2.9%; Score 30.6; DB 4; Length 6693;
Best Local Similarity 53.8%; Pred. No. 23;
Matches 63; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 798 TGTTCACAGGCTTCACAGCTGCTAGACACCTCTTTTACAGTGTTCACAGCGG 857
DB 2197 TGTCCAACTGCTCTTTATCATGTGTGCACTGCTGCTCCACAGACCCCTTTTTCAGAGGGGG 2138

QY 858 CTTTGTATCGGTTTCAGAGCTTACGGTTTCGAATTTATCAAGAGGAGCTGGTTAT 914
DB 2137 CTTTGTCTAGTACCGCTGTATCGTGGGCTGTGTCAGAGTGCCTGTCTTTAT 2081

RESULT 13
US-07-982-712-1
; Sequence 1, Application US/07982712
; Patent No. 5436391
; GENERAL INFORMATION:
; APPLICANT: Hideya FUJIMOTO, Kimiko ITOH
; APPLICANT: Mikihiro YAMAMOTO, and Ko SHIMAMOTO
; TITLE OF INVENTION: Insecticidal protein-encoding Gene, Gramineous
; TITLE OF INVENTION: Plants Transformed with the Gene, and Production Thereof
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Penack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 144 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/982,712
; FILING DATE: 19921127
; CLASSIFICATION: 800
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2172 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; FEATURE:
; NAME/KEY: peptide
; LOCATION: 1 to 2172
; IDENTIFICATION METHOD: E or S
; OTHER INFORMATION:
US-07-982-712-1

Query Match 2.8%; Score 30.4; DB 1; Length 2172;
Best Local Similarity 54.5%; Pred. No. 15;
Matches 61; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 880 ACGGTTCCGAATTTATCAAGACGCTGCTGTTATGGCTCCGAGCTCAAGCTCAGCGGTA 939

DB 821 AGAATTTCCAGCGCAGCTTCCGGGCTCCGCCAGGCGCATCGAGGCGAGCATCAGAGCC 880

QY 940 ACAACGGTAGTTTCATCGTCAATTGCAACAACAACGACGACGACGCTGAGAG 991
DB 881 CGCACCTCATGGACATCTCTCAACAGCATCACCATTACACCGACGCCACAG 932

RESULT 14
US-09-125-619-3/c
; Sequence 3, Application US/09125619
; Patent No. 6437116
; GENERAL INFORMATION:
; APPLICANT: NORRIS, STEVEN J.
; APPLICANT: JING-REN, ZHANG
; APPLICANT: HARDHAM, JOHN M.
; APPLICANT: HOWELL, JERILYN K.
; APPLICANT: BARBOUR, ALAN G.
; APPLICANT: WEINSTOCK, GEORGE M.
; TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
; FILE REFERENCE: US/09/125,619
; CURRENT APPLICATION NUMBER: US/09/125,619
; CURRENT FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 7766
; TYPE: DNA
; ORGANISM: Borrelia burgdorferi
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (127)
; OTHER INFORMATION: R = A OR G
US-09-125-619-3

Query Match 2.8%; Score 30.4; DB 4; Length 7766;
Best Local Similarity 48.3%; Pred. No. 28;
Matches 85; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 542 CCGTTTCAGCTTCCCTCCGGTTTAGTCCAAATGCCACGACGACGATCCAACTCCGC 601
DB 3151 CCTATCAGCAGCCTTCGACGACGACGATATCCAACTTCTCAATTCGACGAGTC 3092

QY 602 AGCTCTGGCATCATCCACTGTGGCTCAGCAACTTTCGCGCAAGAAATGTATCGATGT 661
DB 3091 CCTATAGCCCCCTCAGCTTCCCTTCTCATCTTTCACAGCAAACTTCCATCCT 3032

QY 662 GGCTATTCCATCAACGCAATGATTCGAGCGTCCGAGCTTCTCTTGTATCCA 717
DB 3031 TAGCCATCCCTCAAGCAATAGCAGCAGCAATCTGATCATCTCTTCTATCCA 2976

RESULT 15
US-08-545-528D-1/c
; Sequence 1, Application US/08545528D
; Patent No. 6537773
; GENERAL INFORMATION:
; APPLICANT: Fraser et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragments
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB193P1
; CURRENT APPLICATION NUMBER: US/08/545,528D
; CURRENT FILING DATE: 1995-10-19
; PRIOR APPLICATION NUMBER: US 08/488,018
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/473,545
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 580073
; TYPE: DNA
; ORGANISM: Mycoplasma genitalium

US-08-545-528D-1

Query Match 2.8%; Score 30.4; DB 4; Length 580073;
 Best Local Similarity 47.0%; Pred. No. 2.1e+02;
 Matches 94; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

| | | | |
|----|------|--|------|
| QY | 871 | TCAGACGTTAGCGGTTTCGAATTTATCAGAGCGACGTCGGTTATGGCTCGAGCTCAAGC | 930 |
| DB | 3402 | TCACCTCTTTTCCATTACAGAGAAATCAGGAACAACCTCAACAATTTTCATGCTCTTTT | 3343 |
| QY | 931 | TCAGGCGTAAACAACCGGTAGTTTCATGCTCAATTGCAACAACAACGACGACACGCTGAGA | 990 |
| DB | 3342 | TCAGAGTTACCTTCTTGGACCAATCTCCAATTACCTTTCTCCATCGCTAAAGCTGAGA | 3283 |
| QY | 991 | GACTTCTCCCTAGAGATATACGAGAAACAAGAGCTTCACCGGTTCTATGAGCACCACAACA | 1050 |
| DB | 3282 | AAATACCTTTTATTTTTCACGAAAACCTCAAACTTTAAAGAAGAACTAAAGCGCATTAAACA | 3223 |
| QY | 1051 | GCACGGTCAATCGAACCACTG | 1070 |
| DB | 3222 | ACTGATGCACCAACACCGTG | 3203 |

Search completed: February 2, 2004, 03:00:11
 Job time : 101.543 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 1, 2004, 21:42:55 ; Search time 402.599 Seconds
(without alignments)
9695.297 Million cell updates/sec

Title: US-09-938-842A-1034

Perfect score: 1071

Sequence: 1 atggcgacaattcagaagct.....caggtcgcgaaccactga 1071

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2434939 seqs, 1822278265 residues

Total number of hits satisfying chosen parameters: 4869878

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/FCR_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/FCR06_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US10C_NEW_PUB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US10D_NEW_PUB.seq.*
- 18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| result No. | Score | Match | Length | DB ID | Description |
|------------|-------|-------|--------|-------|---------------------|
| 1 | 1071 | 100.0 | 1071 | 10 | US-09-938-842A-1034 |
| 2 | 1071 | 100.0 | 1071 | 12 | US-09-938-842A-1034 |
| 3 | 448 | 41.8 | 460 | 10 | US-09-924-035A-502 |
| 4 | 439 | 41.0 | 453 | 9 | US-09-770-444-615 |
| 5 | 185 | 17.3 | 185 | 9 | US-09-770-695-257 |
| 6 | 185 | 12.0 | 128 | 11 | US-09-934-455-169 |
| 7 | 129 | 12.0 | 128 | 13 | US-10-225-068-165 |
| 8 | 124.6 | 11.6 | 1231 | 11 | US-09-934-455-137 |
| 9 | 124.6 | 11.6 | 1231 | 13 | US-10-302-267-61 |
| 10 | 124.6 | 11.6 | 1231 | 13 | US-10-225-068-245 |
| 11 | 121.6 | 11.4 | 1604 | 15 | US-10-295-403-147 |
| 12 | 111.2 | 11.0 | 1009 | 11 | US-09-770-961-675 |
| 13 | 111.2 | 10.4 | 1009 | 11 | US-09-934-455-133 |
| 14 | 111.2 | 10.4 | 1009 | 13 | US-10-225-068-135 |
| 15 | 111.2 | 10.4 | 1009 | 15 | US-10-295-403-145 |

| | | | | | | |
|----|-------|------|---------|----|----------------------|----------------------|
| 16 | 111.2 | 10.4 | 1009 | 16 | US-10-278-536-203 | Sequence 203, App |
| 17 | 104.6 | 9.8 | 819 | 12 | US-10-260-238-383 | Sequence 383, App |
| 18 | 100 | 9.3 | 1172 | 11 | US-09-934-455-461 | Sequence 461, App |
| 19 | 98.6 | 9.2 | 1263 | 11 | US-09-934-455-135 | Sequence 135, App |
| 20 | 89 | 8.3 | 955 | 13 | US-10-225-068-181 | Sequence 181, App |
| 21 | 79.6 | 7.4 | 419 | 10 | US-09-878-574-4107 | Sequence 4107, App |
| 22 | 77 | 7.2 | 612 | 12 | US-10-260-238-5623 | Sequence 5623, App |
| 23 | 67.6 | 6.3 | 552 | 12 | US-10-260-238-65 | Sequence 65, Appl |
| 24 | 50 | 4.7 | 228 | 12 | US-10-260-238-5848 | Sequence 5848, App |
| 25 | 49.6 | 4.6 | 297 | 12 | US-10-260-238-5405 | Sequence 5405, App |
| 26 | 43.6 | 4.1 | 447 | 9 | US-09-770-444-761 | Sequence 761, App |
| 27 | 36.6 | 3.4 | 3316 | 15 | US-10-242-056-58 | Sequence 58, Appl |
| 28 | 34 | 3.2 | 654 | 8 | US-08-781-986A-613 | Sequence 613, App |
| 29 | 33.4 | 3.1 | 629 | 13 | US-10-027-633-221892 | Sequence 221892, App |
| 30 | 33.4 | 3.1 | 629 | 14 | US-10-027-633-221892 | Sequence 221892, App |
| 31 | 33.2 | 3.1 | 469 | 9 | US-09-864-761-4365 | Sequence 4365, App |
| 32 | 33.2 | 3.1 | 828 | 15 | US-10-156-761-5087 | Sequence 5087, App |
| 33 | 33.2 | 3.1 | 1893 | 15 | US-10-128-714-7597 | Sequence 7597, App |
| 34 | 33.2 | 3.1 | 9025608 | 15 | US-10-156-761-1 | Sequence 1, Appl |
| 35 | 33 | 3.1 | 671 | 15 | US-10-184-644-346 | Sequence 346, App |
| 36 | 33 | 3.1 | 671 | 15 | US-10-184-634-346 | Sequence 346, App |
| 37 | 32.8 | 3.1 | 1265 | 15 | US-10-198-846-6773 | Sequence 6773, App |
| 38 | 32.6 | 3.0 | 656 | 13 | US-10-027-633-151527 | Sequence 151527, App |
| 39 | 32.6 | 3.0 | 656 | 14 | US-10-027-633-151527 | Sequence 151527, App |
| 40 | 32.6 | 3.0 | 846 | 12 | US-10-320-797-2065 | Sequence 2065, App |
| 41 | 32.6 | 3.0 | 1019 | 12 | US-10-320-797-1065 | Sequence 1065, App |
| 42 | 32.6 | 3.0 | 2320 | 15 | US-10-268-441-5 | Sequence 5, Appl |
| 43 | 32.6 | 3.0 | 3019 | 12 | US-10-320-797-65 | Sequence 65, Appl |
| 44 | 32.4 | 3.0 | 1200 | 10 | US-09-887-576-731 | Sequence 731, App |
| 45 | 32.2 | 3.0 | 429 | 15 | US-10-268-441-3 | Sequence 3, Appl |

ALIGNMENTS

RESULT 1

US-09-938-842A-1034
Sequence 1034, Application US/09938842A
Patent No. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
FILE REFERENCE: SCRIPT1300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,966
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 1034
LENGTH: 1071
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-1034

Query Match 100.0%; Score 1071; DB 10; Length 1071;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1071; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGCGACAAATTCAGAAGCTTCAGAAGTTGAGCGAAGATCAAACTCTAAGACCGGTT 60

Db 1 ATGGCGACAAATTCAGAAGCTTCAGAAGTTGAGCGAAGATCAAACTCTAAGACCGGTT 60

Qy 61 GATCTAACCATCATCAACGGCGTCGAAACGTCGAAACCTTCAAGACCTTCCAAAGTAAT 120

Db 61 GATCTAACCATCATCAACGGCGTCGAAACGTCGAAACCTTCAAGACCTTCCAAAGTAAT 120

121 CCCACAGTGAAGTCTCGAGCCCAAGCCGAGCCGCGTATGCGCGTCTGTTTCAATGTCCTTTA 180
121 CCCACAGTGAAGTCTCGAGCCCAAGCCGAGCCGCGTATGCGCGTCTGTTTCAATGTCCTTTA 180
181 GCTCCACGCTCTTCACAGGACACCAATGAGAGAGCTTCGATTAAGACCGTCCACAG 240
181 GCTCCACGCTCTTCACAGGACACCAATGAGAGAGCTTCGATTAAGACCGTCCACAG 240
241 AAGGTTGAAGGAGAGGAGAGATACGATGCTCCACGCTGCGGCTAGGATTTTT 300
241 AAGGTTGAAGGAGAGGAGAGATACGATGCTCCACGCTGCGGCTAGGATTTTT 300
301 CAATTAATCGAGAGTTAGTCTCAAAATCCGACGCGGAAACGATTCGATGAGTGGAG 360
301 CAATTAATCGAGAGTTAGTCTCAAAATCCGACGCGGAAACGATTCGATGAGTGGAG 360
361 AACGCTGAGCCGCGGATATATAGCCGCGCACCGGATCGGAAACGCTCCGCGCATCGCCATG 420
361 AACGCTGAGCCGCGGATATATAGCCGCGCACCGGATCGGAAACGCTCCGCGCATCGCCATG 420
421 TCGGTTAAGGAACTTAAATATCCGACGACGACGATGCTGATTCGATGAGTGGAA 480
421 TCGGTTAAGGAACTTAAATATCCGACGACGACGATGCTGATTCGATGAGTGGAA 480
481 AATCTGATGAAGAGAAACGTAAGACGCTTCTAACAGTGAATATAGACATAAGCGAC 540
481 AATCTGATGAAGAGAAACGTAAGACGCTTCTAACAGTGAATATAGACATAAGCGAC 540
541 GCCGTTTCAGTCTCCTCGGTTTATAGTCCAAATGCCACGACGACGATCCAACTCCG 600
541 GCCGTTTCAGTCTCCTCGGTTTATAGTCCAAATGCCACGACGACGATCCAACTCCG 600
601 CAAGCTCTGGCATCATCCACTGCTGGCTCAGCACTTCTGCGCAAGGAATGATTCGATG 660
601 CAAGCTCTGGCATCATCCACTGCTGGCTCAGCACTTCTGCGCAAGGAATGATTCGATG 660
661 TGGGCTATTCCATCAAAACGCAATGATTCGACGCTCGGAGCTTCTTCTTGTATCCACAA 720
661 TGGGCTATTCCATCAAAACGCAATGATTCGACGCTCGGAGCTTCTTCTTGTATCCACAA 720
721 ATCGTGTGCTCGTCAATCAGCTCAGTATAGCTTTTCCGCGCGCGCTGCTTCCGCG 780
721 ATCGTGTGCTCGTCAATCAGCTCAGTATAGCTTTTCCGCGCGCGCTGCTTCCGCG 780
781 TCGTCTTACGTCGCGCTGTTTCAACAGCTTCCACGATGCTAGACCACTCTTGTATCAAA 840
781 TCGTCTTACGTCGCGCTGTTTCAACAGCTTCCACGATGCTAGACCACTCTTGTATCAAA 840
841 GTTGTTCACAGCAGCGGCTTTGTATCCGTTTTCAGAGCTTACGCTTGAATTTATCAAGA 900
841 GTTGTTCACAGCAGCGGCTTTGTATCCGTTTTCAGAGCTTACGCTTGAATTTATCAAGA 900
901 GCGAGCTCGGTTATGCTCCGAGCTCAGCTCAGCGCTTAACACCGGTAGTTCATCGTCA 960
901 GCGAGCTCGGTTATGCTCCGAGCTCAGCTCAGCGCTTAACACCGGTAGTTCATCGTCA 960
961 ATTGCAACAACAACGACGACACAGCTGAGAGCTTCTCCCTAGAGATATACGAGAAACAA 1020
961 ATTGCAACAACAACGACGACACAGCTGAGAGCTTCTCCCTAGAGATATACGAGAAACAA 1020
1021 GAGCTTCCAGGTTATGAGACACACACAGCAGCGCTCATCGAACCTGA 1071
1021 GAGCTTCCAGGTTATGAGACACACACAGCAGCGCTCATCGAACCTGA 1071

RESULT 2

US-09-938-842A-1034
; Sequence 1034, Application US/09938842A
; Publication No. US20040009476A9
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel

APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
FILE REFERENCE: SRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 1034
LENGTH: 1071
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-1034

Query Match 100.0%; Score 1071; DB 12; Length 1071;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1071; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCGCAATTCAGAGCTTGAAGAGTTGACGCAAGATCAAACTCTAAGACCGGTT 60
DB 1 ATGGCGCAATTCAGAGCTTGAAGAGTTGACGCAAGATCAAACTCTAAGACCGGTT 60
QY 61 GATCTAACCATCATCAACGGCGTCAGAAACGTCAAACTTCAAGACCTTTCCAAAGTAAAT 120
DB 61 GATCTAACCATCATCAACGGCGTCAGAAACGTCAAACTTCAAGACCTTTCCAAAGTAAAT 120
QY 121 CCCACAGTGAAGTCTCGAGCCCAAGCCGAGCCGCGTATGCGCGTCTGTTTCAATGTCCTTTA 180
DB 121 CCCACAGTGAAGTCTCGAGCCCAAGCCGAGCCGCGTATGCGCGTCTGTTTCAATGTCCTTTA 180
QY 181 GCTCCACGCTCTTCACAGGACACCAATGAGAGAGCTTCGACTTAAAGACCGTCCACAG 240
DB 181 GCTCCACGCTCTTCACAGGACACCAATGAGAGAGCTTCGACTTAAAGACCGTCCACAG 240
QY 241 AAGGTTGAAGGAGAGGAGAGATGAGTACGATGCTGCGACGCTGCGGCTAGGATTTTT 300
DB 241 AAGGTTGAAGGAGAGGAGAGATGAGTACGATGCTGCGACGCTGCGGCTAGGATTTTT 300
QY 301 CAATTAATCTCGAGAGTTAGTCTCAAAATCCGACGCGGCAACGATTCGCTGCTGTTTGGAG 360
DB 301 CAATTAATCTCGAGAGTTAGTCTCAAAATCCGACGCGGCAACGATTCGCTGCTGTTTGGAG 360
QY 361 AACGCTGAGCCGCGGATATAGCCGCGCACCGGTAAGGAGAGCTTCCCGCATCGCCATG 420
DB 361 AACGCTGAGCCGCGGATATAGCCGCGCACCGGTAAGGAGAGCTTCCCGCATCGCCATG 420
QY 421 TCGGTTAAGGAACTTAAATATCCGACGACGACGATGCTGATTCGATGAGTGGAA 480
DB 421 TCGGTTAAGGAACTTAAATATCCGACGACGACGATGCTGATTCGATGAGTGGAA 480
QY 481 AATCTGATGAAGAGAAACGTAAGACGCTTCTAACAGTGAATATAGACATAAGCGAC 540
DB 481 AATCTGATGAAGAGAAACGTAAGACGCTTCTAACAGTGAATATAGACATAAGCGAC 540
QY 541 GCCGTTTCAGTCTCCTCGGTTTATAGTCCAAATGCCACGACGACGATCCAACTCCG 600
DB 541 GCCGTTTCAGTCTCCTCGGTTTATAGTCCAAATGCCACGACGACGATCCAACTCCG 600
QY 601 CAAGCTCTGGCATCATCCACTGCTGGCTCAGCACTTCTGCGCAAGGAATGATTCGATG 660
DB 601 CAAGCTCTGGCATCATCCACTGCTGGCTCAGCACTTCTGCGCAAGGAATGATTCGATG 660
QY 661 TGGGCTATTCCATCAAAACGCAATGATTCGACGCTCGGAGCTTCTTCTTGTATCCACAA 720
DB 661 TGGGCTATTCCATCAAAACGCAATGATTCGACGCTCGGAGCTTCTTCTTGTATCCACAA 720
QY 721 ATCGTGTGCTCGGATCAGCTCAGTATAGCTTTTCCGCGCGCGCTGCTTCCGCG 780

721 ATGCTGGTGGCGCGAATCAGCTCAGTTATAGCTTTTCCGCGCGCGCTGCTTCGCGG 780
781 TCGTCTTACGTCGCCCTGTTCAACAGGCTTCCAGATGGCTAGACCACCTCCTTTACAA 840
781 TCGTCTTACGTCGCCCTGTTCAACAGGCTTCCAGATGGCTAGACCACCTCCTTTACAA 840
841 GTTGTTCCAGCAGCGCGCTTGTATCCGTTTCAGAGTTTACGAGTTTCAATTAACA 900
841 GTTGTTCCAGCAGCGCGCTTGTATCCGTTTCAGAGTTTACGAGTTTCAATTAACA 900
901 GCGAGCTCGGTTATGGCTCCGAGCTCAAGCTCAGCGGTAAACACCGGTAGTTCACTGCA 960
901 GCGAGCTCGGTTATGGCTCCGAGCTCAAGCTCAGCGGTAAACACCGGTAGTTCACTGCA 960
961 ATTGCAACAAACAGCAGCAGCTCAGCTGAGAGACTTCTCCCTAGAGATATACGAGAAACA 1020
961 ATTGCAACAAACAGCAGCAGCTCAGCTGAGAGACTTCTCCCTAGAGATATACGAGAAACA 1020
1021 GAGCTTCCACAGTTTATGAGCAGCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAG 1071
1021 GAGCTTCCACAGTTTATGAGCAGCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAG 1071

RESULT 3

IS-09-924-035A-502/c
Sequence 502, Application US/09924035A
Patent No. US20020142319A1
GENERAL INFORMATION:
APPLICANT: Glach, Jm
TITLE OF INVENTION: Expressed Sequences of Arabidopsis
FILE REFERENCE: 2011US
CURRENT APPLICATION NUMBER: US/09/924,035A
CURRENT FILING DATE: 2000-08-11
PRIOR APPLICATION NUMBER: US 60/148,784
PRIOR FILING DATE: 1999-08-13
NUMBER OF SEQ ID NOS: 900
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 502
LENGTH: 460
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)...(460)
OTHER INFORMATION: n = A,T,C or G
IS-09-924-035A-502

Query Match 41.8%; Score 448; DB 10; Length 460;
Best Local Similarity 99.6%; Pred. No. 8e-145;
Matches 459; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
600 GCAAGCTCGGCATATCCATCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAG 659
460 GCAAGCTCGGCATATCCATCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAG 401
660 GTGGGCTATTCATCAACGCAATGATTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAG 719
400 GTGGGCTATTCATCAACGCAATGATTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAG 341
720 AATCGTGTGCTGCTGATCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAG 779
340 AATCGTGTGCTGCTGATCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAG 282
780 GTGCTCTTACGTCGCGCTGTTCAACAGCTTCCAGCTCAGCTCAGCTCAGCTCAGCTCAG 839
281 GTGCTCTTACGTCGCGCTGTTCAACAGCTTCCAGCTCAGCTCAGCTCAGCTCAGCTCAG 222
840 AGTGTGTTCAACAGCGGCTTGTATCCGTTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 899
221 AGTGTGTTCAACAGCGGCTTGTATCCGTTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 162

900 AGCAGCTCGGTTATGGCTCCGAGCTCAAGCTCAGCGGTAAACACCGGTAGTTCACTGCTC 959
161 AGCAGCTCGGTTATGGCTCCGAGCTCAAGCTCAGCGGTAAACACCGGTAGTTCACTGCTC 102
960 AATTGCAACAAACAGCAGCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAG 1019
101 AATTGCAACAAACAGCAGCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAG 42
1020 AGAGCTTCCACAGTTTATGAGCAGCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAG 1060
41 AGAGCTTCCACAGTTTATGAGCAGCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAG 1

RESULT 4

US-09-770-444-615/c
Sequence 615, Application US/09770444
Patent No. US2002023280A1
GENERAL INFORMATION:
APPLICANT: Gorlach, Jm
APPLICANT: An, Yong-Qiang
APPLICANT: Hamilton, Carol M.
APPLICANT: Price, Jennifer L.
APPLICANT: Raines, Tracy M.
APPLICANT: Yu, Yang
APPLICANT: Rameaka, Joshua G.
APPLICANT: Page, Amy
APPLICANT: Matthew, Abraham V.
APPLICANT: Ledford, Brooke L.
APPLICANT: Woessner, Jeffrey P.
APPLICANT: Haas, William David
APPLICANT: Garcia, Carlos A.
APPLICANT: Krieger, Maja
APPLICANT: Slader, Ted
APPLICANT: Davis, Keith R.
APPLICANT: Allen, Keith
APPLICANT: Hoffman, Neil
APPLICANT: Hurban, Patrick
TITLE OF INVENTION: Expressed Sequences of Arabidopsis
FILE REFERENCE: 2027 (PABA-016PRV)
CURRENT APPLICATION NUMBER: US/09/770,444
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: 60/178,502
PRIOR FILING DATE: 2000-01-27
NUMBER OF SEQ ID NOS: 999
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 615
LENGTH: 453
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)...(453)
OTHER INFORMATION: n = A,T,C or G
US-09-770-444-615

Query Match 41.0%; Score 439; DB 9; Length 453;
Best Local Similarity 99.1%; Pred. No. 1.1e-141;
Matches 450; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
607 CTGGCATCATCCATCTGTGGCTCAGCAACTTCTGCGCGAAGGAATGTATCCGATGTGGGCT 666
453 CTGGCATCATCCATCTGTGGCTCAGCAACTTCTGCGCGAAGGAATGTATCCGATGTGGGCT 394
667 ATTCCATCAACGCAATGATTCGAGCGGTCTGAGAGCTTCTTCTTCTGATCCCAAAATCGCT 726
393 ATTCCATCAACGCAATGATTCGAGCGGTCTGAGAGCTTCTTCTTCTGATCCCAAAATCGCT 334
727 GTTCCGTCGAATCAGCTCAGTTATAGCTTTTCCGCGCGGCTGCTTCCGCGCTGCT 786
333 GTTCCGTCGAATCAGCTCAGTTATAGCTTTTCCGCGCGGCGC-NNNTCCGCGCTGCT 275
787 TAGTGTGCGCTGTTTCAACAGCTTCCAGATGGCTTAGACCACTCTTACAAATGTT 846

Db 274 TACGTCGGCGCTTTTCAACAGCGCTTCCAGATGGGTAGACACCTCCCTTTCAAGATTGTT 215
QY 847 CCAAGCAGCGGCTTTGTATCCGTTTCAAGCTTAGCGGTTTCAATTTATCAAGAGCGACG 906
Db 214 CCAAGCAGCGGCTTTGTATCCGTTTCAAGCTTAGCGGTTTCAATTTATCAAGAGCGACG 155
QY 907 TCGGTTATGGCTCCGAGCTCAAGCTCAGCGGTACCAACCGGTAGTTCATCGTCAATTGCA 966
Db 154 TCGGTTATGGCTCCGAGCTCAAGCTCAGCGGTACCAACCGGTAGTTCATCGTCAATTGCA 95
QY 967 ACAACACGACGACACACGCTCAGAGACTTCTCCCTAGAGATATACGAGAAACAGAGCTT 1026
Db 94 ACAACACGACGACACACGCTCAGAGACTTCTCCCTAGAGATATACGAGAAACAGAGCTT 35
QY 1027 CACAGTTCATGAGACACACACGAGCGGTAT 1060
Db 34 CACAGTTCATGAGACACACACGAGCGGTAT 1

RESULT 5

US-09-770-696-257
; Sequence 257, Application US/09770696
; Patent No. US2001004940A1

GENERAL INFORMATION:

; APPLICANT: Goriach, Jörn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Kricker, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick

; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: 2031US (PARA-020PRV)
; FILE REFERENCE: 2031US (PARA-020PRV)
; CURRENT APPLICATION NUMBER: US/09/770,696
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/178,278
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 911
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 257

; LENGTH: 185

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

US-09-770-696-257

Query Match

Best Local Similarity 17.3%; Score 185; DB 9; Length 185;

Matches 185; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 GCAAGATCAAACTTAAAGACCGTGTGATCAACATCATCAACGCGGTGAGAAAGTCG 94
Db 1 GCAAGATCAAACTTAAAGACCGTGTGATCAACATCATCAACGCGGTGAGAAAGTCG 60
QY 95 AAACCTCAAGACCTTTTCAAGTAATCCACAGTGTCTCGAGCCCAAGCGGAGCGG 154
Db 61 AAACCTCAAGACCTTTTCAAGTAATCCACAGTGTCTCGAGCCCAAGCGGAGCGG 120
QY 155 TGATGCCGCTGCTTTCATGTCTTAGTTCACCGTCTTCACAGGACCACTTGAAGA 214

Db 121 TGATGCCGCTGCTTTCATGTCTTAGTTCACCGTCTTCACAGGACCACTTGAAGA 180
QY 215 GAGCT 219
Db 181 GAGCT 185

RESULT 6

US-09-934-455-169
; Sequence 169, Application US/09934455
; Publication No. US20030121070A1

GENERAL INFORMATION:

; APPLICANT: Adam, Luc
; APPLICANT: Creelman, Robert
; APPLICANT: Dubell, Arnold
; APPLICANT: Heard, Jacqueline
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Keddle, James
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Reuber, Lynne
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Pineda, Omeira

; TITLE OF INVENTION: Genes for Modifying Plant Traits IV

; FILE REFERENCE: MBI-0025

; CURRENT APPLICATION NUMBER: US/09/934,455

; CURRENT FILING DATE: 2001-08-22

; PRIOR APPLICATION NUMBER: 60/227439

; PRIOR FILING DATE: 2000-08-22

; PRIOR APPLICATION NUMBER: MBI-0022

; PRIOR FILING DATE: 2001-11-16

; PRIOR APPLICATION NUMBER: MBI-0023

; PRIOR FILING DATE: 2001-04-17

; NUMBER OF SEQ ID NOS: 516

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 169

; LENGTH: 1728

; TYPE: DNA

; ORGANISM: Arabidopsis thaliana

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (106)...(1575)

; OTHER INFORMATION: G1064

US-09-934-455-169

Query Match

Best Local Similarity 12.0%; Score 129; DB 11; Length 1728;

Matches 172; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 189 GTCTTCGACAGACCACTTGAAGAGAGCTTCGACTAAAGACCGTCACACGAGGTGA 248
Db 420 GCGAGCTAAAAGCCACCGTTGAAACGAGCGTCGACGAAAGACGACACGAAAGTAGA 479
QY 249 AGGAAGAGGAGAGATACGATCGCTGCCACGCTGCGGTAGGATTTTCAATTAAC 308
Db 480 CGAAGAGGAGAGAGATACGATCGCTGCCACGCTGCGGTAGGATTTTCAATTAAC 539
QY 309 TCGAGAGTGTAGTCACAAATCCGACGCGGAAACGATTCGGTGTGTGGAGAACGCTGA 368
Db 540 GCGAGAGTGTAGTCACAAATCCGACGCGGTAGACAAATAGAGTGGCTTCTTCAACAAAGCTGA 599
QY 369 GCGGCGGATTATAGCCGACGCGGTACGGGAAACGCTTCGCGCATCCGATGCGGTAA 428
Db 600 ACCATCTGTAAATCGCGCCGACCGGAAACGCGGAAACAAATCCCGGGAATTTCACTTCTTTAA 659
QY 429 C 429
Db 660 C 660

RESULT 7

US-10-225-068-165

Sequence 165, Application US/10225068
Publication No. US20030217383A1

GENERAL INFORMATION:

APPLICANT: Mendel Biotechnology, Inc.

APPLICANT: Reuber, T. Lynne

APPLICANT: Riechmann, Jose Luis

APPLICANT: Heard, Jacqueline E.

APPLICANT: Jiang, Cai-Zhong

APPLICANT: Adam, Luc J.

APPLICANT: Dubell, Arnold T.

APPLICANT: Ratcliffe, Oliver

APPLICANT: Pineda, Omaisra

APPLICANT: Yu, Guo-Liang

APPLICANT: Broun, Pierre E.

TITLE OF INVENTION: STRESS-RELATED POLYNUCLEOTIDES AND

TITLE OF INVENTION: POLYPEPTIDES IN PLANTS

FILE REFERENCE: 51442002040

CURRENT APPLICATION NUMBER: US/10/225,068

PRIOR FILING DATE: 2002-08-09

PRIOR APPLICATION NUMBER: 60/310,847

PRIOR FILING DATE: 2001-08-09

PRIOR APPLICATION NUMBER: 60/336,049

PRIOR FILING DATE: 2001-11-19

PRIOR APPLICATION NUMBER: 60/338,692

PRIOR FILING DATE: 2001-12-11

PRIOR APPLICATION NUMBER: 10/171,468

PRIOR FILING DATE: 2002-06-14

NUMBER OF SEQ ID NOS: 246

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 165

LENGTH: 1728

TYPE: DNA

ORGANISM: Arabidopsis thaliana

FEATURE:

NAME/KEY: CDS

LOCATION: (106)... (1575)

US-10-225-068-165

Query Match 12.0%; Score 129; DB 13; Length 1728;
Best Local Similarity 71.0%; Pred. No. 2.4e-33;
Matches 171; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

189 GTCTTCGACGAGCACCACCATTAAGAGAGCTTCGACTAAGACCTCACAAGAGTTGA 248

420 GCGAGTAAAGGCCACCTGTGAACGAGCTCGACGAAAGACCCACACAGAAAGTAGA 479

249 AGAAGAGGAGAGATACGATCGCTCCACGCTGCGGCTAGATTTTCAATTAAC 308

480 CGAAGAGGAGAGATTAAGATGCGGCGTTATGTGACGCTAGGGTTTTCAGCTAAC 539

309 TCAGAGTTAGGTCAAAATCCGAGCGCGGAAACGATTGCGTGTGTTGGAGAACGCTGA 368

540 GCGAGAGTAGTCTATAATCCGAGCGTGAGACAATAGATGGCTTCTTCAACAGCTGA 599

369 GCGGCGGATATAGCGGCACGCGGTACGGGAACGGTTCCGCCCATCGCATGTCGGTTAA 428

600 ACCATCTGTAATCGCGCGCACCGGAACCGGAACAATCCCGGGAATTCACCTTCTTTAAA 659

429 C 429

560 C 560

RESULT 8

US-09-934-455-137

Sequence 137, Application US/09934455

Publication No. US20030121070A1

GENERAL INFORMATION:

APPLICANT: Adam, Luc

APPLICANT: Creelman, Robert

APPLICANT: Dubell, Arnold

APPLICANT: Heard, Jacqueline

APPLICANT: Jiang, Cai-Zhong

APPLICANT: Keddle, James

APPLICANT: Pilgrim, Marsha

APPLICANT: Ratcliffe, Oliver

APPLICANT: Reuber, Lynne

APPLICANT: Riechmann, Jose Luis

APPLICANT: Yu, Guo-Liang

APPLICANT: Pineda, Omaisra

APPLICANT: Broun, Pierre E.

APPLICANT: Jiang, Cai-Zhong

APPLICANT: Adam, Luc

APPLICANT: Reuber, Lynne

APPLICANT: Zhang, James

APPLICANT: Yu, Guo-Liang

APPLICANT: Jiang, Cai-Zhong

APPLICANT: Samaha, Raymond

APPLICANT: Pilgrim, Marsha

APPLICANT: Creelman, Robert

TITLE OF INVENTION: Genes for Modifying Plant Traits IV

FILE REFERENCE: MBI-0025

CURRENT APPLICATION NUMBER: US/09/934,455

PRIOR FILING DATE: 2001-08-22

PRIOR APPLICATION NUMBER: 60/227439

PRIOR FILING DATE: 2000-08-22

PRIOR APPLICATION NUMBER: MBI-0022

PRIOR FILING DATE: 2001-11-16

PRIOR APPLICATION NUMBER: MBI-0023

PRIOR FILING DATE: 2001-04-17

NUMBER OF SEQ ID NOS: 516

SOFTWARE: PatentIn version 3.1

SEQ ID NO 137

LENGTH: 1231

TYPE: DNA

ORGANISM: Arabidopsis thaliana

FEATURE:

NAME/KEY: CDS

LOCATION: (107)... (1084)

OTHER INFORMATION: G805

US-09-934-455-137

Query Match 11.6%; Score 124.6; DB 11; Length 1231;

Best Local Similarity 65.0%; Pred. No. 6.5e-32;

Matches 184; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

Qy 202 CCACCATTTGAGAGAGCTTCGACTAAGACCTGACACGAGGTTGAGGAGAGGAGA 261

Db 239 CCTCCTCTAAACGAACTCTACTAAGACCGACACACCAAAAGTCGAGGCGCGCT 298

Qy 262 AGGATACGGATGCTGCCACGCTGGGGCTAGATTTTCAATTAACTCGAGAGTTAGGT 321

Db 299 CGATCGGTATGCTGCCATGTGTGTCGACGTCTTTCAGCTCACAGTGAGTTGGT 358

Qy 322 CACAAATCCGACGGCGGAAACGATTCGGTGTGTTGGAGAACGCTGAGCGCGGATTATA 381

Db 359 CACAAATCCGACGGCGGAAACGATTCGGTGTGTTGGAGAACGCTGAGCGCGGATTATA 418

Qy 382 GCGCGCACGGGTACGGGAAACGCTTCGGCGCATCGCTGCTGTTAAACGGAACCTTAAA 441

Db 419 GCGCTACAGGAGTGGAAACCATTCGGCTAACTCACTTTTAAACATCTCCTTGGT 478

Qy 442 ATCCGACGACGACGAGACGCTGATTCGATGATGGGTGAAATC 484

Db 479 AGCTCAAGATCTTCTCTCTGCTGCTCATCTTGGTACAACTC 521

RESULT 9

US-10-302-267-61

Sequence 61, Application US/10302267

Publication No. US20030229915A1

GENERAL INFORMATION:

APPLICANT: Keddle, James

APPLICANT: Fromm, Michael

APPLICANT: Heard, Jacqueline

APPLICANT: Riechmann, Jose Luis

APPLICANT: Adam, Luc

APPLICANT: Broun, Pierre

APPLICANT: Pineda, Omaisra

APPLICANT: Reuber, Lynne

APPLICANT: Zhang, James

APPLICANT: Yu, Guo-Liang

APPLICANT: Jiang, Cai-Zhong

APPLICANT: Samaha, Raymond

APPLICANT: Pilgrim, Marsha

APPLICANT: Creelman, Robert

; TITLE OF INVENTION: PLANT GENE SEQUENCES II
; FILE REFERENCE: MBI-0007
; CURRENT APPLICATION NUMBER: US/10/302,267
; CURRENT FILING DATE: 2002-11-22
; PRIOR APPLICATION NUMBER: US/09/506,720
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/120,880
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: 60/121,037
; PRIOR FILING DATE: 1999-02-22
; PRIOR APPLICATION NUMBER: 60/124,278
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: 60/129,450
; PRIOR FILING DATE: 1999-04-15
; PRIOR APPLICATION NUMBER: 60/135,134
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: 60/144,153
; PRIOR FILING DATE: 1999-07-15
; PRIOR APPLICATION NUMBER: 60/161,143
; PRIOR FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: 60/162,656
; PRIOR FILING DATE: 1999-11-01
; NUMBER OF SEQ ID NOS: 218
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 61
; LENGTH: 1231
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (107)..(1081)
; OTHER INFORMATION: G805
US-10-302-267-61

Query Match 11.6%; Score 124.6; DB 12; Length 1231;
Best Local Similarity 65.0%; Pred. No. 6.5e-32;
Matches 184; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 202 CCACCATTTGAAGAGAGCTTCGACTAAAGACCGTGCACACGAGGTTGAAGGAGGAGGAGA 261
DB 239 CCTCCTCTTAACGAACTCTACTTAAGACCGACACACCAACGTCGAGGCCGCGGT 298

QY 262 AGGATACGATGCTGCGACGCTGTCGCGGTAGGATTTTCAATTAACCTCAGAGTTAGGT 321
DB 299 CGGATCCGCTATGCTGCGCATGTGCTGCAGTGCTTTTCAGCTCACAACGTGAGCTTGGT 358

QY 322 CACAAATCCGACGCGGAAACGATTCCGCTGTTGTTGGAGAACGTCGACCGCGGATTATA 381
DB 359 CACAAATCCGACGCTGAACTATTGAGTGGCTACTCCACAGCAGNACCGCGGTTATA 418

QY 382 GCGGCCACGGGTACGGGAACGGTTCCCGCCATGCCATGTCGGTTAACGGAACCTTAAAA 441
DB 419 GCGGCTACAGGACTGGAACCAATTCGGCTAACTTCTCTTAAACATCTCACTTCGT 478

QY 442 ATCCCGACGACGACGCTGATTCTGATATGGTGAAATC 484
DB 479 AGCTCAAGATCTTCTCTCTGCTGCTCATCTTCGTACAACTC 521

RESULT 10
US-10-225-068-245
; Sequence 245 Application US/10225068
; Publication No. US20030217383A1
; GENERAL INFORMATION:
; APPLICANT: Mendel Biotechnology, Inc.
; APPLICANT: Reuber, T. Lynne
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Heard, Jacqueline E.
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Adam, Luc J.
; APPLICANT: Dubell, Arnold T.
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Pineda, Omaira

; APPLICANT: Yu, Guo-Liang
; APPLICANT: Broun, Pierre E.
; TITLE OF INVENTION: STRESS-RELATED POLYNUCLEOTIDES AND
; TITLE OF INVENTION: POLYPEPTIDES IN PLANTS
; FILE REFERENCE: 51442002040
; CURRENT APPLICATION NUMBER: US/10/225,068
; CURRENT FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: 60/310,847
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/336,049
; PRIOR FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: 60/338,692
; PRIOR FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: 10/171,468
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 246
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 245
; LENGTH: 1231
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (107)..(1084)
US-10-225-068-245

Query Match 11.6%; Score 124.6; DB 13; Length 1231;
Best Local Similarity 65.0%; Pred. No. 6.5e-32;
Matches 184; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 202 CCACCATTTGAAGAGAGCTTCGACTAAAGACCGTGCACACGAGGTTGAAGGAGGAGGAGA 261
DB 239 CCTCCTCTTAACGAACTCTACTTAAGACCGACACACCAACGTCGAGGCCGCGGT 298

QY 262 AGGATACGATGCTGCGACGCTGTCGCGGTAGGATTTTCAATTAACCTCAGAGTTAGGT 321
DB 299 CGGATCCGCTATGCTGCGCATGTGCTGCAGTGCTTTTCAGCTCACAACGTGAGCTTGGT 358

QY 322 CACAAATCCGACGCGGAAACGATTCCGCTGTTGTTGGAGAACGTCGACCGCGGATTATA 381
DB 359 CACAAATCCGACGCTGAACTATTGAGTGGCTACTCCACAGCAGNACCGCGGTTATA 418

QY 382 GCGGCCACGGGTACGGGAACGGTTCCCGCCATGCCATGTCGGTTAACGGAACCTTAAAA 441
DB 419 GCGGCTACAGGACTGGAACCAATTCGGCTAACTTCTCTTAAACATCTCACTTCGT 478

QY 442 ATCCCGACGACGACGCTGATTCTGATATGGTGAAATC 484
DB 479 AGCTCAAGATCTTCTCTCTGCTGCTCATCTTCGTACAACTC 521

RESULT 11
US-10-295-403-147
; Sequence 147 Application US/10295403
; Publication No. US20030101481A1
; GENERAL INFORMATION:
; APPLICANT: Heard, Jacqueline
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Adam, Luc
; APPLICANT: Broun, Pierre
; APPLICANT: Pineda, Omaira
; APPLICANT: Reuber, Lynne
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Keddle, James
; APPLICANT: Benito, Maria-Ines
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Fromm, Mike
; TITLE OF INVENTION: PLANT GENE SEQUENCES I
; FILE REFERENCE: MBI-0003
; CURRENT APPLICATION NUMBER: US/10/295,403
; CURRENT FILING DATE: 2002-11-15
; PRIOR APPLICATION NUMBER: US/09/394,519

;; PRIOR FILING DATE: 1999-09-13
;; PRIOR APPLICATION NUMBER: 60/101,349
;; PRIOR FILING DATE: 1998-09-22
;; PRIOR APPLICATION NUMBER: 60/103,312
;; PRIOR FILING DATE: 1998-10-06
;; PRIOR APPLICATION NUMBER: 60/108,734
;; PRIOR FILING DATE: 1998-11-17
;; PRIOR APPLICATION NUMBER: 60/113,409
;; PRIOR FILING DATE: 1998-12-22
;; NUMBER OF SEQ ID NOS: 170
;; SOFTWARE: Patent In Ver. 2.0
;; SEQ ID NO 147
;; LENGTH: 1604
;; TYPE: DNA
;; ORGANISM: Arabidopsis thaliana
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (143)..(1345)
;; OTHER INFORMATION: G802
JS-10-295-403-147

Query Match 11.4%; Score 121.6; DB 15; Length 1604;
Best Local Similarity 64.6%; Pred. No. 8.5e-31;
Matches 181; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
Y 188 CGTCTTCGACAGACACCACTTGAAGAGAGCTTCGACTAAAGACCGTCACACGAGGTTG 247
b 285 CATCTACGGACAGCGGACTACTACAGAGGGTTCGACTAAAGACCGTCACACGAGGTCG 344
Y 248 AAGAAAGAGGAGAGGATACGATCCCTGCCACGCTGTCGCGCTAGGATTTTCAATTAA 307
b 345 ACGAAAGAGACGTCGATTCGATCGCGCGCTATGTGCAGCTAGAGTTTTTCAAGTTAA 404
Y 308 CTCGAGAGTGTGTACAACTCGACGGCGAAGACGATCGGTGGTGTGGAGACGCTG 367
b 405 CGAGAGGTTAGGTCTAATAATCGATGGTGAACACTATTGAATGGCTCTTCAACAGCTG 464
Y 368 AGCCGGCGATTATAGCGGCACCGGTACGGGAACGGTTCGCCCATCGCATGCGGTTA 427
b 465 AGCCAGCTATTGTGTCTGTACAGGTACAGGAATTTCCGGCGAATCTCTCTATTAA 524
Y 428 ACGAACTTAAATCCGACGACGACGACGACGCTGATTC 467
b 525 GTGTTTCGTACGAAGTAGTGGATCGATCTCTCTGCTCC 564

RESULT 12

US-09-770-961-675
; Sequence 675, Application US/09770961
; Publication No. US20030115639A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jörn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Kricker, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: thaliana
; FILE REFERENCE: 2026 (PARA-015PRV)

;; CURRENT APPLICATION NUMBER: US/09/770,961
;; CURRENT FILING DATE: 2001-01-26
;; PRIOR APPLICATION NUMBER: 60/178,466
;; PRIOR FILING DATE: 2000-01-27
;; NUMBER OF SEQ ID NOS: 999
;; SOFTWARE: PabSeq for Windows Version 4.0
;; SEQ ID NO 675
;; LENGTH: 490
;; TYPE: DNA
;; ORGANISM: Arabidopsis thaliana
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: (1)..(490)
;; OTHER INFORMATION: n = A,T,C or G
US-09-770-961-675

Query Match 11.0%; Score 118; DB 11; Length 490;
Best Local Similarity 73.3%; Pred. No. 7.2e-30;
Matches 151; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
Q 211 AAGAGAGCTTCGACTAAAGACCGTCACACGAGGTTGAAGGAGAGGAGAGGATACGG 270
b 249 AAGAGAGAGCTCTAACAAAGACAGACACACTAAAGTCGAAGGTAGAGGTCCAGCAATTCGG 308
Q 271 ATGCTCTCCACGCTGTCGGCTAGGATTTTCAATTAACTCGAGAGTTAGGTCACAATCC 330
b 309 ATGCTCTCTTTGTGCTGCTAGGATTTTCAATTCCTAGAGATTTGGTCTATAATCT 368
Q 331 GACGGCGAAACGATTCGGTGGTTGTTGGAGAACGCTGAGCCGCGGANTATAGCCGCCAG 390
b 369 GATGCTGAACACTATCCAGTGGCTGCTTCAACAGCTGAGCCATCGATTATTGCAGCTACT 428
Q 391 GGTACGGGACGGTTCGCCCATCGC 416
b 429 GGTTCAGAACTATACCGGCTCTGCG 454

RESULT 13

US-09-934-455-133
; Sequence 133, Application US/09934455
; Publication No. US20030121070A1
; GENERAL INFORMATION:
; APPLICANT: Adam, Luc
; APPLICANT: Creelman, Robert
; APPLICANT: Dubell, Arnold
; APPLICANT: Heard, Jacqueline
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Keddle, James
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Reuber, Lynne
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Pineda, Omaira
; TITLE OF INVENTION: Genes for Modifying Plant Traits IV
; FILE REFERENCE: MBI-0025
; CURRENT APPLICATION NUMBER: US/09/934,455
; CURRENT FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/227439
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: MBI-0022
; PRIOR FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: MBI-0023
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 516
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 133
; LENGTH: 1009
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (27)..(746)

OTHER INFORMATION: G801
US-09-934-455-133

Query Match 10.4%; Score 111.2; DB 11; Length 1009;
Best Local Similarity 71.6%; Pred. No. 2.7e-27;
Matches 146; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
QY 209 TGAAGAGAGCTTCGACTAAACACCGTACACAGAGGTTGAAGGAGGAGGAGGATAC 268
DB 103 TTAAGAAACACCGCTTAAGATCGACACAGCAAGTTGACGGAAGGAGGAGGATTC 162
QY 269 GGATGCTGCCACGTGTGCGGTAGGATTTTCAATTAACTCGAGAGTTAGGTACAAAT 328
DB 163 GTATGCCAATCATTTTGGCAGCTCGAGTTTTTCAATTGACCAGAGAGTTAGGTACAAAGT 222
QY 329 CCGACGGCGAAACGATTCGGTGGTTTGGAGAGACGCTGAGCCGCGCATATAGCCGCCA 388
DB 223 CCGATGGTCAACACCATAGAGTGGCTTCTCCGTCAAGCTGAGCCTTCTATCATAGCCGCCA 282
QY 389 CGGGTACGGGAACGGTTCGCCGCCA 412
DB 283 CTGGAACCTGGCACTACTCCGGCGA 306

RESULT 14

US-10-225-068-135
; Sequence 135, Application US/10225068
; Publication No. US20030217383A1
; GENERAL INFORMATION:
; APPLICANT: Mendel Biotechnology, Inc.
; APPLICANT: Reuber, T. Lynne
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Heard, Jacqueline E.
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Adam, Luc J.
; APPLICANT: Dubell, Arnold T.
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Pineda, Omalra
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Broun, Pierre E.
; TITLE OF INVENTION: STRESS-RELATED POLYPEPTIDES IN PLANTS
; FILE REFERENCE: 51442002040
; CURRENT APPLICATION NUMBER: US/10/225,068
; CURRENT FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: 60/310,847
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/336,049
; PRIOR FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: 60/338,692
; PRIOR FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: 10/171,468
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 246
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 135
; LENGTH: 1009
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (27)...(746)
US-10-225-068-135

Query Match 10.4%; Score 111.2; DB 13; Length 1009;
Best Local Similarity 71.6%; Pred. No. 2.7e-27;
Matches 146; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
QY 209 TGAAGAGAGCTTCGACTAAACACCGTACACAGAGGTTGAAGGAGGAGGAGGATAC 268
DB 103 TTAAGAAACACCGCTTAAGATCGACACAGCAAGTTGACGGAAGGAGGAGGATTC 162
QY 269 GGATGCTGCCACGTGTGCGGTAGGATTTTCAATTAACTCGAGAGTTAGGTACAAAT 328

DB 163 GTATGCCAATCATTTTGGCAGCTCGAGTTTTTCAATTGACCAGAGAGTTAGGTACAAAGT 222
QY 329 CCGACGGCGAAACGATTCGGTGGTTTGGAGAGCGCTGAGCCGCGCATATAGCCGCCA 388
DB 223 CCGATGGTCAACACCATAGAGTGGCTTCTCCGTCAAGCTGAGCCTTCTATCATAGCCGCCA 282
QY 389 CGGGTACGGGAACGGTTCGCCGCCA 412
DB 283 CTGGAACCTGGCACTACTCCGGCGA 306

RESULT 15

US-10-295-403-145
; Sequence 145, Application US/10295403
; Publication No. US20030101481A1
; GENERAL INFORMATION:
; APPLICANT: Heard, Jacqueline
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Adam, Luc
; APPLICANT: Broun, Pierre
; APPLICANT: Pineda, Omalra
; APPLICANT: Reuber, Lynne
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Keddle, James
; APPLICANT: Zhang, James
; APPLICANT: Benito, Maria-Ines
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Fromm, Mike
; TITLE OF INVENTION: PLANT GENE SEQUENCES I
; FILE REFERENCE: MBI-0003
; CURRENT APPLICATION NUMBER: US/10/295,403
; CURRENT FILING DATE: 2002-11-15
; PRIOR APPLICATION NUMBER: US/09/394,519
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: 60/101,349
; PRIOR FILING DATE: 1998-09-22
; PRIOR APPLICATION NUMBER: 60/103,312
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: 60/108,734
; PRIOR FILING DATE: 1998-11-17
; PRIOR APPLICATION NUMBER: 60/113,409
; PRIOR FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 145
; LENGTH: 1009
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (27)...(743)
; OTHER INFORMATION: G801
US-10-295-403-145

Query Match 10.4%; Score 111.2; DB 15; Length 1009;
Best Local Similarity 71.6%; Pred. No. 2.7e-27;
Matches 146; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
QY 209 TGAAGAGAGCTTCGACTAAACACCGTACACAGAGGTTGAAGGAGGAGGAGGATAC 268
DB 103 TTAAGAAACACCGCTTAAGATCGACACAGCAAGTTGACGGAAGGAGGAGGATTC 162
QY 269 GGATGCTGCCACGTGTGCGGTAGGATTTTCAATTAACTCGAGAGTTAGGTACAAAT 328
DB 163 GTATGCCAATCATTTGCGCAGCTCGAGTTTTTCAATTGACCAGAGAGTTAGGTACAAAGT 222
QY 329 CCGACGGCGAAACGATTCGGTGGTTTGGAGAGACGCTGAGCCGCGCATATAGCCGCCA 388
DB 223 CCGATGGTCAACACCATAGAGTGGCTTCTCCGTCAAGCTGAGCCTTCTATCATAGCCGCCA 282
QY 389 CGGGTACGGGAACGGTTCGCCGCCA 412

Db 283 CTGGAACTGGCACTACTCCGGCGA 306

Search completed: February 2, 2004, 01:10:23
Job time : 414.599 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

M nucleic - nucleic search, using sw model

run on: February 1, 2004, 21:43:46 ; Search time 2477.95 Seconds
(without alignments)
10504.671 Million cell updates/sec

Title: US-09-938-842A-1034

Perfect score: 1071
Sequence: 1 atgggacaaatcagaagct.....caggtcatcgaaaccactga 1071

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

EST:

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pin:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_nam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 409.6 | 38.2 | 798 | 28 | BH511920 BOHGA14TF |
| 2 | 382 | 35.7 | 486 | 9 | AV825375 AV825375 |
| 3 | 360.8 | 33.7 | 373 | 10 | BES22768 M28F8STM |
| 4 | 314 | 29.3 | 335 | 14 | T88453 12149 Lambd |

| | | | | | | | | |
|---|----|-------|------|-----|----|----------|------------|-----------|
| C | 5 | 311.2 | 29.1 | 656 | 28 | BH450567 | BH450567 | BOGD041TR |
| | 6 | 292.6 | 27.3 | 388 | 14 | 235068 | ATT83707 | GI |
| | 7 | 255.2 | 23.8 | 378 | 10 | BES23004 | M31B6STM | |
| | 8 | 195.2 | 18.2 | 725 | 28 | BH543331 | BOGNC02TF | |
| C | 9 | 165 | 15.4 | 680 | 28 | BH470366 | BHC8133TF | |
| | 10 | 164.4 | 15.4 | 791 | 14 | CB292965 | AD292965 | AI0 |
| | 11 | 157.8 | 14.7 | 634 | 9 | AU239448 | CU239448_0 | |
| | 12 | 154.6 | 14.4 | 652 | 9 | AW697442 | UF239448 | |
| C | 13 | 149.2 | 13.9 | 439 | 14 | CA929907 | NF118G06S | |
| | 14 | 146.4 | 13.7 | 747 | 13 | BQ865644 | QTU2CA.P4 | |
| | 15 | 145.8 | 13.6 | 649 | 10 | BF070046 | QGC5H15.Y | |
| C | 16 | 145.2 | 13.6 | 459 | 9 | AI973663 | st63b02.Y | |
| | 17 | 132.4 | 12.4 | 716 | 12 | AI973663 | sd08a06.Y | |
| C | 18 | 132 | 12.3 | 673 | 13 | BM437189 | VVA015P11 | |
| | 19 | 130.6 | 12.2 | 626 | 10 | BQ874162 | QG74J06.Y | |
| | 20 | 130 | 12.1 | 661 | 10 | BE440597 | SP47e10.Y | |
| | 21 | 129.4 | 12.1 | 586 | 28 | BF278295 | GA_EB003 | |
| C | 22 | 129 | 12.0 | 514 | 9 | BH741132 | GT37g10.G | |
| | 23 | 129 | 12.0 | 556 | 12 | AV442079 | AV442079 | |
| | 24 | 129 | 12.0 | 568 | 12 | BM522413 | saX98f11. | |
| | 25 | 127.8 | 11.9 | 736 | 29 | BQ403313 | CA_EB005 | |
| | 26 | 127.8 | 11.9 | 773 | 28 | BZ460479 | BONJW58TR | |
| | 27 | 127.4 | 11.9 | 671 | 9 | BH866140 | BOMX456TR | |
| | 28 | 126.6 | 11.8 | 522 | 9 | AW649873 | EST328327 | |
| | 29 | 126.6 | 11.8 | 649 | 9 | AJ503095 | AJ503095 | |
| | 30 | 126.6 | 11.8 | 657 | 9 | AW695867 | NF100A09S | |
| | 31 | 126 | 11.8 | 907 | 10 | AW695783 | NF098G07S | |
| | 32 | 125.6 | 11.7 | 693 | 13 | BG442740 | GA_EB001 | |
| | 33 | 124.4 | 11.6 | 652 | 10 | BQ405430 | GA_EB008 | |
| | 34 | 123.6 | 11.5 | 728 | 29 | BF269051 | GA_EB000 | |
| | 35 | 123.6 | 11.5 | 775 | 29 | BZ519530 | BOMQX23TR | |
| | 36 | 123.6 | 11.5 | 785 | 28 | CB972449 | CAB330001 | |
| | 37 | 123.2 | 11.5 | 416 | 10 | BH664697 | BOHUX48TR | |
| | 38 | 123 | 11.5 | 830 | 10 | BF052551 | EST437781 | |
| | 39 | 122.6 | 11.4 | 422 | 9 | BG447197 | GA_EB004 | |
| | 40 | 122.6 | 11.4 | 527 | 12 | AJ489125 | AJ489125 | |
| | 41 | 122.6 | 11.4 | 562 | 10 | BI932752 | EST552641 | |
| | 42 | 121.6 | 11.4 | 581 | 10 | BE463339 | EST534562 | |
| | 43 | 121.2 | 11.3 | 815 | 10 | BG446467 | GA_EB003 | |
| | 44 | 121 | 11.3 | 519 | 12 | BG445804 | GA_EB002 | |
| | 45 | 121 | 11.3 | 622 | 10 | BM111907 | EST559443 | |
| | 46 | 121 | 11.3 | 622 | 10 | BE344238 | EST409400 | |

ALIGNMENTS

RESULT 1
BH511920/c
LOCUS BH511920
DEFINITION BH511920 BOHG Brassica oleracea genomic clone BOHGA14, genomic survey sequence.
ACCESSION BH511920
VERSION BH511920.1 GI:17720010
KEYWORDS GSS
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Rukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
REFERENCE 1 (bases 1 to 798)
AUTHORS Town.C.D., Van Aken,S., Utterback.T., Koo,H. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished
COMMENT Other GSSs: BOHGA14TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: ctown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.

```
FEATURES             Location/Qualifiers
source               1..798
                    /organism="Brassica oleracea"
                    /mol_type="genomic DNA"
                    /strain="TOL000DH3"
                    /db_xref="taxon:3712"
                    /clone_lib="BOHGA14"
                    /note="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT          161 a 210 c 225 g 202 t
ORIGIN
Query Match          38.2%; Score 409.6; DB 28; Length 798;
Best Local Similarity 80.3%; Pred. No. 1.9e-118;
Matches 525; Conservative 0; Mismatches 84; Indels 45; Gaps 2;
QY 184 CCACGCTCTTCACAGGACCACTTGAAGAGAGCTTCGACTTAAGACCGTCACACGAG 243
DB 798 CCGTCGTCTCAATAGGACCGCAATGAAGAGAGCTTCGACTTAAGACCGTCACACGAG 739
QY 244 GTTGAAGGAGGAGGAGAGATACGGATGCTGCCAGCTGTGGCGCTAGGATTTTCAA 303
DB 738 GTTGAAGGAGGAGGAGGAGATACGGATGCTGCCAGCTGTGGCGCTAGGATTTTCCAG 679
QY 304 TTAATCTGAGAGTTAGTTCACAAATCCGACGCGCAAGATTCGGTGTGTTGGAGAAC 363
DB 678 TTGACTCGAGAGCTTGGTCACAAATCCGACGCGGAGAGATTCGGTGTGTTGGAGAAC 619
QY 364 GCTGAGCGCGCGGATATATAGCCGCCACGGTACGGGAACGGTTCGCCCATCGCATGTGCG 423
DB 618 GCGGAGCGCGCGGATATATAGCCGCCACGGTACGGGAACGGTTCGCCCATCGCATGTGCG 559
QY 424 GTTAAAGGAACTTTAAATATCCGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 483
DB 558 GTCAACGGAACGCTAAAATATCCGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 499
QY 484 CTGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 543
DB 498 CCGGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 439
QY 544 GTTTCAGCTTCTCCGGTTTACGCTTCAATTCGACGAGAGAGAGAGAGAGAGAGAGAGAG 603
DB 438 GTCTCCGTTTCTCCGGTTTACGCTTCAATTCGACGAGAGAGAGAGAGAGAGAGAGAGAG 411
QY 604 GCTCTGGCATATCCACTGTGCTGAGCAACTTCTACAGTGAATATAGATATAGAGAGAGAG 663
DB 410 -----CATCCACCGCGGCTCAACAAACGCTGCGCGAGAGAGATGATTCGGATGTGG 361
QY 664 GCTATTCATCAACAGCAATGATTCGACGAGTTCGAGAGTTCCTTCTGATTCACAAATC 723
DB 360 GCGTTTCCATCAACAGCAATGATTCGACGAGTTCGAGAGTTCCTTCTGATTCACAAACGTA 301
QY 724 GCTGTTCCTGCAATCAGCTCAGTTATAGTTTTC---CGCGCGCGCTGCTTCGCGG 780
DB 300 GTGCGCGCGTGAACAGAGCTCAGATATAGTTTATCCGCGCGCGCTGCTGCTTACCG 241
QY 781 TCGTCTTACGCTCGCGCTGCTTCAACAGGCTTCCAGAGTTCGAGAGTTCAGTCCCT 834
DB 240 TCGAGCTATGTCGCGCGCTTTCAGAGGCTTCTTCTATGCTAGACCACTTCT 187

RESULT 2
AV825375
LOCUS              486 bp mRNA linear EST 01-APR-2002
DEFINITION         AV825375 RAFL7 Arabidopsis thaliana cDNA clone RAFL07-08-P04 5',
                    mRNA sequence.
ACCESSION          AV825375
VERSION            AV825375.1 GI:19867435
KEYWORDS           Arabidopsis thaliana (thale cress)
SOURCE             Arabidopsis thaliana
ORGANISM           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

REFERENCE
AUTHORS            Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,
                    Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,
                    Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.,
                    and Shinozaki, K.
TITLE              Large scale analysis of Arabidopsis full-length cDNA (2002b)
JOURNAL            Unpublished
COMMENT            Contact: Motoaki Seki
                    Plant Functional Genomics Research Group
                    RIKEN Genomic Sciences Center
                    3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
                    Tel: 81-298-36-4359
                    Fax: 81-298-36-9060
                    Email: mseki@tc.riken.go.jp
                    An Arabidopsis full-length cDNA library was constructed essentially
                    as reported previously (Seki et al., 1998) cDNA cleaved with BamHI
                    and XhoI was ligated to modified Lambda PLC-1 vector (Carninci et
                    al., submitted for publication) digested with BamHI and SalI. This
                    clone is in a modified pBluescript vector. Please visit our web
                    site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further
                    details.
FEATURES             Location/Qualifiers
source               1..486
                    /organism="Arabidopsis thaliana"
                    /mol_type="mRNA"
                    /db_xref="taxon:3702"
                    /clone="RAFL07-08-P04"
                    /dev_stage="rosette plants"
                    /lab_host="DH10B"
                    /clone_lib="RAFL7"
                    /note="Site 1: BamHI; Site 2: SalI; subjected to
cold-treated (1, 2, 5, 10, 24 hr)"
BASE COUNT          131 a 118 c 117 g 117 t 3 others
ORIGIN
Query Match          35.7%; Score 382; DB 9; Length 486;
Best Local Similarity 99.2%; Pred. No. 8.4e-110;
Matches 382; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGGGACAAATTCAGAAAGTTGAAGAAGTTGAGCAAGATCAAACTCTAAGAGCCGTT 60
DB 102 ATGGGACAAATTCAGAAAGTTGAAGAAGTTGAGCAAGATCAAACTCTAAGAGCCGTT 161
QY 61 GATCTAACCATCATCAACGGCGTCAGAACGTCGAACTTCAGACCTTCCAGTAAT 120
DB 162 GATCTAACCATCATCAACGGCGTCAGAACGTCGAACTTCAGACCTTCCAGTAAT 221
QY 121 CCCACAGTGAAGTCTGAGCCCAAGCGGAGCGGTCGTCCTGCGTCGTTTCAATGCTTTA 180
DB 222 CCCACAGTGAAGTCTGAGCCCAAGCGGAGCGGTCGTCCTGCGTCGTTTCAATGCTTTA 281
QY 181 GCTCCACCGTCTTCGACGAGCACCATTTGAAGAGAGCTTCGACTTAAGACCGTCACAG 240
DB 282 GCTCCACCGTCTTCGACGAGCACCATTTGAAGAGAGCTTCGACTTAAGACCGTCACAG 341
QY 241 AAGGTTGAAGGAGAGGAGAGAGATACGATGCTGCGTCGTCGCTAGGATTTT 300
DB 342 AAGGTTGAAGGAGAGGAGAGAGATACGATGCTGCGTCGTCGCTAGGATTTT 401
QY 301 CAATTAACTCGAGAGTTAGTTCACAAATCCGACGCGGAAACGATTCGGTGGTGTGGAG 360
DB 402 CAATTAACTCGAGAGTTAGTTCACAAATCCGACGCGGAAACGATTCGGTGGTGTGGAG 461
QY 361 AACGCTGAGCGCGGATTTATAGCCG 385
DB 462 AACGCTGAGCGCGGATTTATAGCCG 486

RESULT 3
BE522768
LOCUS              373 bp mRNA linear EST 19-MAR-2001
```



```

Db      121 CACAGTGTCTCGAGCCGAGCGGACGTNNATGCGTCNTTTTCAATGTTTATG 180
QY      183 TCCACCGTCTTCGACAGACCAACATTGAAGAGAGCTTCGACTAAAGACCGTCAACGAA 242
Db      181 TCCACCGTNTTCGACAGGACCAACATTGAAGAGAGCTTCGACTAAAGACCGTCAACGAA 240
QY      243 GGTTCAGAGAGAGGAGGAGGATGATCGGATCGCTGCCACGCTGCGGCTAGGATTTTCA 302
Db      241 GGTTCAGAGAGAGGAGGAGGATGATCGGATCGCTGNNACGTTGCGGCTAGGATTTTCA 300
QY      303 ATTAATCTCGAGAGTTAGGTTCACAAATCCGACGGCG 337
Db      301 ATTAATCTCGAGGTTAGGTTCACAAATCCGCGGG 335

RESULT 5
BH450567/c
LOCUS      BH450567
DEFINITION BOGDQ41TR BOGD Brassica oleracea genomic clone BOGDQ41, genomic survey sequence.
ACCESSION BH450567
VERSION   BH450567.1
KEYWORDS  GSS.
SOURCE    Brassica oleracea
ORGANISM  Brassica oleracea
REFERENCE 1 (bases 1 to 656)
AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE     Whole genome shotgun sequencing of Brassica oleracea
JOURNAL   Unpublished
COMMENT   Other GSSs: BOGDQ41TF
          Contact: Chris Town
          TIGR
          9712 Medical Center Drive, Rockville, MD 20850, USA.
          Tel: 301-838-3523
          Fax: 301-838-0208
          Email: cdtown@tigr.org
          DNA is from a doubled haploid provided by Tom Osborn.
          Seq primer: TR
          Class: sheared ends.
FEATURES   Location/Qualifiers
            1..656
               /organism="Brassica oleracea"
               /mol_type="genomic DNA"
               /strain="T0100DH3"
               /db_xref="taxon:3712"
               /clone="BOGDQ41"
               /clone_lib="BOGD"
               /note="Vector: pBOS1; Site_1: BstXI; 2-3 kb sheared genomic DNA inserted into pBOS1 using BstXI linkers"
BASE COUNT 142 a 167 c 190 g 157 t
ORIGIN
Query Match      29.1%; Score 311.2; DB 28; Length 656;
Best Local Similarity 78.4%; Pred. No. 3.2e-87;
Matches 431; Conservative 0; Mismatches 73; Indels 46; Gaps 3;

QY      288 GGTAGGATTTTCAATTAATCTCGAGAGTTAGGTTCACAAATCCGACGCGGAAAGATTTCG 347
Db      656 GGTAGGATTTTCAATTAATCTCGAGAGTTAGGTTCACAAATCCGACGCGGAGAGATTTCG 597
QY      348 GTGGTGTGAGAGACGTCGAGCGCGGATATATAGCCGCCACGGGTACGGAAACGGTTCC 407
Db      596 GTGGTGTG-TGGAGAACGGCGGACGGCGGATATATCGCCGCCACGGGTACGGAAACGGTTCC 538
QY      408 CCCCATCGCATGTGCGTTTAAACGAACTTAAATATCCGACGACGACGACGCTGATTC 467
Db      537 CCCCATCGCATGTGCGTTCACGAAACGCTTAAATATCCGACGACGACGACGCTGATTC 478
QY      468 TGATATGGGTGAAATCTGATGAGAGAAACGTTAAACGACCTTCTTACAGTGTAGTATAT 527

```

```

Db      477 TGACGTGGCGGAAAATCCGGGAGAGAGACGTAAGGACCTTCTACCACTAGTATAT 418
QY      528 AGACATAAGCGAGCGGTTTTCAGTTCTCCGTTTACCTCCAAATTCGCCAGCAGCAAC 587
Db      417 AGACATAAGCAAGACGGTCTCCGTTTCTCCGTTTACCTCCGTTTACCTCCGTTTACCTCC 376
QY      588 GATCCAACTCCGCAAGCTCTGGCATCATCCACTGTGGCTCAGCAACTTCTGCGCGCAAGG 647
Db      375 -----GGCATCCACCGCGGCTCAACAACGCTGCGCGCAGG 340
QY      648 AATGATTCGATGTGGGTATTTCCATCAACGCAATGATTCGACGCTTCGAGGTTTCTT 707
Db      339 GATGATTCGATGTGGGTATTTCCATCAACGCAATGATTCGACGCTTCGAGGTTTCTT 280
QY      708 CTTGATTCACAAATCGTGTGGTCCGATCAATCAAGCTCAGTATATAGCTTTC---CGGC 764
Db      279 CTTGTTCAAAACGATGTGGCGCGCTCGAACGAGCTCAGATATAGCTTATCCGCGCGC 220
QY      765 CCGCGTGTCTTCGCGCTGTCTTACGTCCGCGCTGTTTCAACAGGCTTCCACGATGGCTAG 824
Db      219 CGTGTCTGTTTACCGTCGACGTATGTGCGCGCTTTCAGCAGGCTTCTTCTATGGCTAG 160
QY      825 ACCACCTCT 834
Db      159 ACCACCTCT 150

RESULT 6
Z35068
LOCUS      Z35068
DEFINITION ATTS3707 Gif-Seeda Arabidopsis thaliana cDNA clone YAI344 5', mRNA sequence.
ACCESSION Z35068
VERSION   Z35068.1
KEYWORDS  EST.
SOURCE    Arabidopsis thaliana (thale cress)
ORGANISM  Arabidopsis thaliana
REFERENCE 1 (bases 1 to 388)
AUTHORS   CNRS
TITLE     The Arabidopsis thaliana transcribed genome: the GDR cDNA program
JOURNAL   Unpublished
COMMENT   Contact: Berthomieu P., Guerrier D., Giraudat J.
          Genetique Moleculaire d'Arabidopsis
          ISV - UPR40, CNRS
          Avenue de la Terrasse, 1198 Gif-sur-Yvette Cedex, France
          Email: Giraudat@cnrs-gif.fr.
FEATURES   Location/Qualifiers
            1..388
               /organism="Arabidopsis thaliana"
               /mol_type="mRNA"
               /db_xref="taxon:3702"
               /clone="YAY344"
               /clone_lib="Gif-Seeda"
               /note="Vector: Lambda ZAPII non-oriented; Physiological condition: greenhouse plants. tissue type: Green siliques"
BASE COUNT 102 a 94 c 108 g 83 t 1 others
ORIGIN
Query Match      27.3%; Score 292.6; DB 14; Length 388;
Best Local Similarity 97.8%; Pred. No. 1.9e-81;
Matches 317; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

QY      86 GAAAGCTCGAAATCTCAAGACCTTTCGAAGTAAATCCACAGTGAAGTCTCGAGCCCAAGG 145
Db      1 GAAAGCTCGAAATCTCAAGACCTTTCGAAGTAAATCCACAGTGAAGTCTCGAGCCCAAGG 60
QY      146 CCGAGCCGCTGATGCGGCTGTTTCAATGTTTCTAGCTCCACCGCTTTCGACGAGCCAC 205
Db      61 CCGAG-CGTGAATGCGGCTGTTTCAATGTTTCTAGCTCCACCGCTTTCGACGAGCCAC 119

```



```

Db      145  ACGGTCTCCGTTTCTCCGCTTATAGTC----- 173
QY      600  GCAAGCTCTGGCATCATCCACTGTGGCTCAGCAACTCTGCCGCAAGAAATGATCCGAT 659
Db      174  -----CGGATTCACCGCGCTCACAAACGCTGCCGACGGGATTCGAT 222
QY      660  GTGGCTATTCATCAAAACGCAATGATTCGACGGTGGAGCTTCTTCTTGATTCACCA 719
Db      223  GTGGCGGTTTCATCAAAACGCAAGTGTTCGACGGTGGAGCTTCTTCTTGTTCAAAA 282
QY      720  AATCGCTGTGCTCGATCAGCTCAGCTTATAGCTTTTC---CCGCGCGCTGCTTC 776
Db      283  CGTACTGCGCGCTCGAACAGCTCAGATATAGCTATCCGCGCGCGCTGCTGCTTT 342
QY      777  GCCGCTGCTTACGTCGCGCTGTGTTCACAGCTTCCACGATGCTAGACCACTCT 834
Db      343  ACCGTGACGATGTGCGCGCGCTTCAGCAGCTTCTCTATGCTAGACCACTCT 400

RESULT 9
LOCUS      BH470366                680 bp      DNA      linear      GSS 13-DEC-2001
DEFINITION BOHC133TF BOHC Brassica oleracea genomic clone BOHC133, genomic
survey sequence.
ACCESSION  BH470366
VERSION    BH470366.1  GI:17678477
KEYWORDS   GSS.
SOURCE     Brassica oleracea
ORGANISM   Brassica oleracea
REFERENCE  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Brassica.
AUTHORS    Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
TITLE      Whole genome shotgun sequencing of Brassica oleracea
JOURNAL    Unpublished
COMMENT    Other GSSs: BOHC133TR
Contact:   Chris Town
TIGR       9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
FEATURES   Location/Qualifiers
            1..680
            /organism="Brassica oleracea"
            /mol_type="genomic DNA"
            /strain="TO1000DH3"
            /db_xref="taxon:3712"
            /clone="BOHC133"
            /clone_lib="BOHC"
            /note="Vector: PHOS1; Site 1: BstXI; 2-3 kb sheared
            genomic DNA inserted into PHOS1 using BstXI linkers"
BASE COUNT 145 a 188 c 175 g 172 t
ORIGIN
Query Match 15.4%; Score 165; DB 28; Length 680;
Best Local Similarity 72.5%; Pred.No. 9.5e-41;
Matches 263; Conservative 0; Mismatches 55; Indels 45; Gaps 2;

QY      475  GGTGAATCTGATGAAGAAGAAACGTAACGACTTCTTAACAGTGAAGTATATAGACATA 534
Db      668  GGCGGAATCCGGCGAAGAGAGAGAGTAAGGACCTTCTACAGTGAATATAGACATA 609
QY      535  ACGGACGCGTTTCAGCTTCTCCCGTTTACCTCCAAATGCGACGACCAACGATCAA 594
Db      608  AACGAACCGGCTCCGTTCTCCCGGTTTAGTCTCGG----- 572
QY      595  CTTCCGCAAGCTCTGGCATCATCCACTGTGGCTCAGCAACTTCTGCCGCAAGGATGAT 654

```

```

Db      571  -----CATCACCGCGCTCAACAAACGCTCCGCGCAGGATGATT 531
QY      655  CCGATGTGGCTATTCCATCAAAACGCAATGATTCGACGGTCGGAGCTTCTTCTTGATT 714
Db      530  CCGATGTGGCGTTTCCATCAAAACGCAAGTGTTCGACGGTTGGAGCTTCTTCTTGTT 471
QY      715  CCACAAATCGCTGTGCTCGTGAATCAGCTCAGTTATTAGCTTTTC---CGCGCGCGCT 771
Db      470  CAAACGCTAGCTGGCGCTCGAAACGAGCTCAGATATTAGCTTATCCGCGCGCGCTGCT 411
QY      772  GCTTCGCGCTGTCTTACGTGCGCTGTTCACAGGCTTCCAGATGCTAGACCACT 831
Db      410  GTTTACGCTGACGTATGTCCGCGCTTCAGCAGGCTTCTTCTATGCTAGACCACT 351

832 CCT 834
350 CCT 348

RESULT 10
LOCUS      CB292965                791 bp      mRNA      linear      EST 28-FEB-2003
DEFINITION UCRC01_05bd03_g1 Washington Navel orange cold acclimated flavedo &
albedo cDNA library Citrus sinensis cDNA clone UCRC01_05bd03, mRNA
sequence.
ACCESSION  CB292965
VERSION    CB292965.1  GI:28618422
KEYWORDS   EST.
SOURCE     Citrus sinensis
ORGANISM   Citrus sinensis
REFERENCE  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Sapindales; Rutaceae; Citrus.
AUTHORS    Close, R.J., Collura, K., Fenton, R.D., Feuerbacher, O., Kim, H.R.,
Kudrna, D., Wamamaker, S., Wing, R. and Yu, Y.
TITLE      Development of EST Resources and New Genetic Markers for California
Citrus
JOURNAL    Unpublished
COMMENT    Contact: Timothy Close
Department of Botany & Plant Sciences, University of California
Riverside, CA, 92521-0124
Tel: 9097873318
Fax: 9097874437
Email: timothy.close@ucr.edu
Seq primer: T3.
FEATURES   Location/Qualifiers
            1..791
            /organism="Citrus sinensis"
            /mol_type="mRNA"
            /cultivar="Washington Navel"
            /db_xref="taxon:2711"
            /clone="UCRC01_05bd03"
            /tissue_type="Rind containing flavedo and albedo"
            /dev_stage="Mature fruit"
            /lab_host="E. coli TUC121"
            /clone_lib="Washington Navel orange cold acclimated
            flavedo & albedo cDNA library"
            /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
            Site 1: EcoRI; Site 2: XhoI; Plants were grown in the
            field at University of California, Riverside Agricultural
            Operations since 1983. The scion was Washington Navel
            orange and the rootstock Carizzo Citrange. Tissue from
            mature fruit was harvested at mid-day in January 2002
            during a cold spell, when pre-dawn temperatures were
            approximately -2 to 2 degrees C. Approximately 2 cm median
            sections of the rind were excised in the field from
            several fruits, then wrapped in aluminum foil and frozen
            quickly in dry ice. Total RNA was extracted using a phenol
            extraction procedure described in J. Japanese Soc. Hort.
            Sci. (1996) 64:809-814. Poly(A) RNA was purified, a cDNA
            library was made, and 1 million primary lambda cDNA clones
            were in vivo excised to give a population of pBluescript

```

SK(-) phagemids. All steps to this point were performed in the TU Close lab at the University of California, Riverside (Fenton). Phagemids were plated, plasmid DNA purified, cDNA clones archived, and DNA sequences determined bi-directionally using an ABI3530 at the Arizona Genomics Institute, University of Arizona (Collura, Feuerbacher, Kim, Kudrna, Wing, Yu). Chromatogram files were transmitted to UC Riverside (by Yu), then processed at UC Riverside (by Wanmaker) using the Harvest pipeline (<http://harvest.ucr.edu>) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a phred 17 region of at least 100 bases were deposited to GenBank."

```

BASE COUNT      203 a 230 c 192 g 166 t
ORIGIN
Query Match      15.4%; Score 164.4; DB 14; Length 791;
Best Local Similarity 71.2%; Pred. No. 1.6e-40;
Matches 232; Conservative 0; Mismatches 91; Indels 3; Gaps 1;

2y 214 AGAGCTTCGACTAAACACCTCAGACAGGTTGAGGAGGAGGAGGAGGATGATG 273
    |||||
2b 56 AGAGCGCTTCCAAAGATCGGCACCAAGGTCGAAGCGCGCGGTAGGATTAGAATG 115
    |||||
2y 274 CCTGCCACGCTGTGCGGTAGGATTTTCAATTAACCTCGAGAGTTAGGTCACAAAATCCGAC 333
    |||||
2b 116 CCGCCACAGCTGTGCGCCAGGATCTTTCACCTAACCCGCGAATCGGTCAAGTCCGAC 175
    |||||
2y 334 GCGCAACAGTTCGGTGTGTTGGAGAACGCTGACCGCGGATATAGCCGCCAGGCT 393
    |||||
2b 176 GGTGAACACCATCCGTTGGTTACTGGAACACGCGGACGCCAGCCATTATTGAGGCCACTGCG 235
    |||||
2y 394 ACGGGAACGGTTCGCCCATCGCATGTGCTGTTAAACGGAACCTTAAATATCCGACGAG 453
    |||||
2b 236 ACTGGCACGGTCCCGCTATCGCGTTCGGTTAACGGAACGTTAAAGATTCCCAACAG 295
    |||||
2y 454 ACGAACGCTGATTCTGATATGGTGTGAATCTGATGAAGAGAAACGTTAAACGCTTCT 513
    |||||
2b 296 TCTCCAGCGAAACCTG---CCGCGAAGAACTGCCGAGGAAACGCGGAAGAGACCGTGT 352
    |||||
2y 514 AACAGTGAGTATATACATAGCGA 539
    |||||
2b 353 AACAGCGAGTTCATTGATTTAACGA 378
    |||||

```

```

RESULT 11
LOCUS      AU239448      634 bp  mRNA  linear  EST 01-APR-2002
DEFINITION AU239448 RAFL19 Arabidopsis thaliana cDNA clone RAFL19-81-E01 5',
            mRNA sequence.
ACCESSION  AU239448
VERSION    AU239448.1 GI:19878617
KEYWORDS   EST.
SOURCE     Arabidopsis thaliana (thale cress)
ORGANISM   Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE  1 (bases 1 to 634)
AUTHORS   Seki,M., Nurusaka,M., Ishida,J., Kamiya,A., Satou,M., Nakajima,M.,
            Akiyama,K., Enju,A., Oono,Y., Sakurai,T., Carninci,P., Kawai,J.,
            Itoh,M., Ishii,Y., Arakawa,T., Shibata,K., Shinagawa,A., Muramatsu
            M., Hayashizaki,Y. and Shinozaki,K.
TITLE     Large scale analysis of Arabidopsis full-length cDNA
JOURNAL   Unpublished
COMMENT   Contact: Motoaki Seki
            Plant Functional Genomics Research Group
            RIKEN Genomic Sciences Center
            2-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: 81-298-36-4359
            Fax: 81-298-36-9060
            Email: mseki@rtc.riken.go.jp
            An Arabidopsis full-length cDNA library was constructed essentially

```

as reported previously (Seki et al., 1998). cDNA cleaved with BamHI and XhoI was ligated to modified Lambda FIC-1 vector (Carninci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified plasmid vector. Please visit our web site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further details.

```

FEATURES             Location/Qualifiers
     source            1..634
                        /organism="Arabidopsis thaliana"
                        /mol_type="mRNA"
                        /db_xref="taxon:3702"
                        /clone="RAFL19-81-E01"
                        /tissue_type="mixture of silique and flower"
                        /lab_host="DH10B"
                        /clone_lib="RAFL19"
BASE COUNT      187 a 137 c 174 g 133 t
ORIGIN
Query Match      14.7%; Score 157.8; DB 9; Length 634;
Best Local Similarity 75.9%; Pred. No. 1.8e-38;
Matches 195; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 195 GACAGGACCAACCTTGAAGAGAGCTTCGACTAAAGACCGTCAACGAGAGTTGAAGGAAG 254
    |||||
Db 187 GACTAAACCGGCTCCGAGAGAGACCGACTTCTAAAGACCGTCAACGAGAGTAGAGGACG 246
    |||||
QY 255 AGGAGAGAGATAGGATGCTCCAGGTGTCGGCTAGGATTTTCAATTAACTCGAGA 314
    |||||
Db 247 AGTCCGAGGATCCGAATCCGCGGCTGCTGCTCGGTCTTTCAATTAACCGCTGA 306
    |||||
QY 315 GTTAGGTCACAAAATCCGACGCGGCGAAACGATTCCGTGTTGTTGGAGAACGCTGAGCCGCG 374
    |||||
Db 307 ACTTGTCACAAATCCGACGCGGAGAACGATACGTTGTTATTGGAACGAGCTGAACCGCG 366
    |||||
QY 375 GATTATAGCCGCCACGGGTACGGGAACGGTTCCGCCCATCGCCCATCGTGGTTAAACGGAAC 434
    |||||
Db 367 GATAATTGAAGCAACCGGAACCGGAACCTGTACCGGCTATTGCTGTATCGTTTAACGGAAC 426
    |||||
QY 435 CTTAAAAATCCCGACGA 451
    |||||
Db 427 TTAAAAATCCCGACGA 443
    |||||

```

```

RESULT 12
LOCUS      AW697442      652 bp  mRNA  linear  EST 15-JUN-2000
DEFINITION AW697442 Developing stem Medicago truncatula cDNA clone
            NF11806ST 5', mRNA sequence.
ACCESSION  AW697442
VERSION    AW697442.1 GI:7572204
KEYWORDS   EST.
SOURCE     Medicago truncatula (barrel medic)
ORGANISM   Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.
REFERENCE  1 (bases 1 to 652)
AUTHORS   He,X.-Z., Shadle,G., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell
            C.J., Flores,H.R., Inman,J.T., Weller,J.W., May,G.D. and Dixon
            R.A.
TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation
            Medicago truncatula stem library
JOURNAL   Unpublished
COMMENT   Contact: Dixon RA
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7302
            Fax: 580 221 7380
            Email: radixon@noble.org
            Insert Length: 652 Std Error: 0.00

```

Plate: 118 row: G column: 06
Seq primer: TCACACAGGAACAGCTATGAC.
Location/Qualifiers
1. .652

FEATURES
source
/organism="Medicago truncatula"
/mol_type="mRNA"
/db_xref="taxon:3880"
/clone="NFI18G06ST"
/tissue_type="stem"
/dev_stage="Pooled developmental"
/clone_lib="Developing stem"
/note="Vector: Lambda Zap; Contains a mixture of
intermodal stem segments"
BASE COUNT 221 a 190 C 119 G 122 T
ORIGIN

Query Match 14.4%; Score 154.6; DB 9; Length 652;
Best Local Similarity 72.9%; Pred. No. 1.9e-37;
Matches 199; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
QY 180 AGCTCCACCGTCTTCGACAGGACCCACCATTTGAAGAGAGCTTCGACTAAAGACCGTCACAC 239
Db 228 AGCTCAGCTCAGCCCAACAAATTCAAAAGAGATCATCTACAAAGACCGTCACAC 287
QY 240 GAAGTTTGAAGAGAGAGAGAGATACGATGCTCCGACGTCGCGCTAGGATTTT 299
Db 288 AAAAGTCGAGGTCGAGCGCGAGAGATCCGAATGCTCGCCACATGTCGAGTCGATCTT 347
QY 300 TCAATTAATCGAGAGTTAGTCACAAATCCGAGCGGAGAAACGATTCGGTGTGTTGGA 359
Db 348 TCAGTCCACCGGAGAACTCGGACATAGTCCGATGCGGAAACAAATACGTCGGCTCCGCA 407
QY 360 GAAGCTGTAGCGCGCGGATTTAGCGCGCCAGCGGTACGGGAAACGTTCCCGCCATGCCAT 419
Db 408 ACAGCGGAGCCAGCTATCATTCGCGCCACCGGTACCGGACATCCGACCATGCCAT 467
QY 420 GTCGTTTAAAGGAACTTAAATTCGCGAGC 452
Db 468 GTCAGTCAACGGAACACTAAATAATCCCCACCAC 500

RESULT 13
CA929907/c
LOCUS
DEFINITION MTU2CA.P4.C10 Aspen apex cDNA Library Populus tremuloides cDNA,
mRNA sequence.
ACCESSION CA929907
VERSION CA929907.1 GI:27418388
KEYWORDS EST
SOURCE Populus tremuloides (quaking aspen)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids I; Malpighiales; Salicaceae; Populus.
REFERENCE 1 (bases 1 to 439)
AUTHORS Ranjan, P., Kao, Y.-Y., Harding, S.A., Jiang, H., Joshi, C.P. and Tsai
C.-J.
TITLE Expressed sequence tags from Aspen
JOURNAL Unpublished
COMMENT Contact: Tsai C-J
Plant Biotech Research Center
Michigan Technological University, School of Forest Resources &
Environmental Science
1400 Townsend Drive, Houghton, MI 49931-1295, USA
Tel: 906 487 2914
Fax: 906 487 2915
Email: chtsai@mtu.edu.
BASE COUNT 221 a 190 C 119 G 122 T
ORIGIN

FEATURES
source
1. .439
/organism="Populus tremuloides"
/mol_type="mRNA"
/db_xref="taxon:3693"
/clone_lib="Aspen apex cDNA Library"

BASE COUNT 91 a 98 C 126 G 124 T
ORIGIN
Query Match 13.9%; Score 149.2; DB 14; Length 439;
Best Local Similarity 62.0%; Pred. No. 7.9e-36;
Matches 254; Conservative 0; Mismatches 153; Indels 3; Gaps 1;
QY 145 CGGAGCGCGTGATGCGCGTCTTTCAATGTCTTTAGCTCCACCGTCTTCGACAGGACCA 204
Db 435 GCTATGCAGATGCAGACGCCAATGCCAATTTCAATGCTCTGCCAGTCACTACACAGCA 376
QY 205 CCATTGAAGAGAGCTTCGACTAAAGACGTCACACGAGGTTGAAGAGAGGAGGAGG 254
Db 375 ACAACCGGACCGCATCCACCAAGACCGCCACACAAAGTGGAGGCGCGTGGCCGTAGG 316
QY 265 ATAGCGATGCTGCGCCAGCTGTGGCGGTAGGATTTTCAATTAATCGAGAGTTAGGTAC 324
Db 315 ATCCGAATACCGGTACTTGTGCTGCGCCGATCTTCCAATTGACCGGAGATTAGGCCAT 256
QY 325 AATCCGACGCGGAAAGATTCGGTGTGTTGAGAGAACGCTGAGCGCGGCGATTATAGCC 384
Db 255 AATCAGACGCGTGAACCGGTAGATGCTCTTTGAACATCTGAGGAAGCTATTATTGAA 196
QY 385 GCCACGGGTACGGGAACGCTTCCGCGCATCGCCATGCGGTAAACGGAACCTTAAAAATC 444
Db 195 GCAACCGGACCGGACCGTCCCTGCTATCGCGTTCGCGTGTGGTGGACTCTCAAAATC 136
QY 445 CCGACGACGAGAAACGCTGATTCGATATGGGTGAAATCTGATGAAGAGAAACGTAA 504
Db 135 CCTCAACAAC--CTCAACAATAGTAATTCATTAACGAAACGCGGAAACCGGAAG 79
QY 505 CGACCTTCAACGAGTATATAGACATAAGCAGCGCGTTCAGCTTC 554
Db 78 CGCCTTCGAATTCGATTCGCGATATTAGTGAAGACGACCCCGTTTC 29

RESULT 14
BQ865644
LOCUS
DEFINITION BQ865644.YG.ab1 QG_ABCD1 lettuce salinas Lactuca sativa cDNA clone
Q865644, mRNA sequence.
ACCESSION BQ865644
VERSION BQ865644.1 GI:22251099
KEYWORDS EST
SOURCE Lactuca sativa
ORGANISM
Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
Cichorieae; Lactuca.
REFERENCE 1 (bases 1 to 747)
AUTHORS Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
Lin, H., van Damme, M., Lavelle, D., Chevallier, P., Ziegler, J., Ellison
P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z.,
Church, S., Jackson, L. and Bradford, K.
TITLE Lettuce and sunflower ESTs from the Compositae Genome Project
JOURNAL Unpublished
COMMENT Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Amundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.edu [michelmore@ucdavis.edu]
singleton, see http://cgdb.ucdavis.edu/ for details.
Plate: Q865 row: H column: 15.
FEATURES
source
1. .747
/organism="Lactuca sativa"
/mol_type="mRNA"
/cultivar="Salinas"

```

/db xref="taxon:4236"
/cclone="QC5H15"
/lab host="E.coli"
/cclone_lib="QC_ABCDI lettuce salinas"
/notes="Vector: pBRCN5A1AB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgdb.ucdavis.edu/TAG LIB=QC ABCDI lettuce salinas
TAG TISSUE=flowers pre-fertilized
TAG SEQ=CGTTGACGGG"
BASE COUNT 199 a 179 c 199 g 168 t 2 others
ORIGIN

Query Match 13.7%; Score 146.4; DB 13; Length 747;
Best Local Similarity 70.4%; Pred. No. 8.2e-35;
Matches 195; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

2Y 224 CTAAGACCGCTCACAGAGTTGAAGGAGAGGAGAGGATACGATGCTGCCACGT 283
Db 338 CTAAGACCGCCACACAGAGTGGAGCGCGTGGCAGGAGAAATCCGATGCCAGCGGTGT 397
2Y 284 GTCCGGCTAGGATTTTCAATTAACTCGAGAGTTAGTTCACAAATCCGACGGCGAAGCA 343
Db 398 GCGCTGCTAGAGCTTCCAGTTTACCTCGGAACTCGTTCACAGTACAGCGCGAGAGA 457
2Y 344 TTCCGGTGTGTTGGAGAACGCTGACCGCGGATATACGCCCGGATGACGGGAACGG 403
Db 458 TTCCGGTGTGTTGGAGAACGCGGAGCGGCTATCATTTGAAGCCACCGGAGCGGACTG 517
2Y 404 TTCCCGCCATCGCATGTCGTTTACCGAACCTTAAATCCGACGACGACGACGCTG 463
Db 518 TTCCCGCAATCGCGTTANTTTAACGGAGCTGTGAAATTTCCGACGACATCAAGTAACG 577
2Y 464 ATTCTGATATGGTGAAATCTGATGAAGAAGAAACG 500
Db 578 GCAAGGAATGGAGATGCTAGGAGAGCGGAAG 614

```

```

RESULT 15
3F070046
LOCUS
DEFINITION
PROTEIN ; mRNA sequence.
BF070046 649 bp mRNA linear EST 06-DEC-2001
Gm-c1053-1227 5' similar to TR:064647 064647 PUTATIVE DNA BINDING
PROTEIN ; mRNA sequence.
BF070046 1 GI:10847341
EST.
Glycine max (soybean)
Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 649)
Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna
,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
,R., Waterston,R. and Wilson,R.
Public Soybean EST Project
Unpublished
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800

```

```

Fax: 314 286 1810
Email: estowatson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntsville, AL 35801 For further information
call: (800)-533-4363 or contact via email: ccu@resgen.com
Insert Length: 1706 Std Error: 0.00
High quality sequence stop: 415.
Location/Qualifiers
1. 649
/organism="Glycine max"
/mol type="mRNA"
/db xref="taxon:3847"
/cclone="GENOME SYSTEMS CLONE ID: Gm-c1053-1227"
/tissue_type="Whole seedling, 3 week old, greenhouse
grown"
/lab host="DH10B"
/cclone_lib="Gm-c1053"
/notes="Vector: pBluescript II SK+; Site 1: EcoRI; Site_2:
XhoI; The Harosoy NIL was constructed and seed was
provided by Dr. J. Specht, University of Nebraska
(Shoemaker and Specht, 1995). The cDNA library was
constructed from mRNA isolated from whole seedlings of 3
week old greenhouse grown plants. Complementary DNA was
synthesized from mRNA using a primer consisting of a
poly(dT) sequence with a XhoI restriction site and a 3'
anchor. EcoRI adapters were ligated to the blunt-ended
cDNA fragments followed by XhoI digestion. The cDNA
fragments were directionally cloned into the EcoRI-XhoI
restriction site of the pBluescript vector. The ligated
cDNA fragments were transformed into DH10B host cells
(GibcoBRL). This library was constructed in cooperation
with Dr. Paul Keim's laboratory at Northern Arizona
University."
BASE COUNT 158 a 162 c 170 g 159 t
ORIGIN

Query Match 13.6%; Score 145.8; DB 10; Length 649;
Best Local Similarity 66.2%; Pred. No. 1.2e-34;
Matches 210; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 136 GAGCCCAAGCGGAGCGGTGATGCCGTGTTTCAATGCTTTAGTCCACCGTCTCG 195
Db 329 GTGCACGTGCCGCGGGGATGCTATGTCCATGTCAGGCTCTGCCCCAGGCCAAGCC 388
QY 196 ACAGGACCACTTGAAGAGAGCTTCGACTAAGACCGTCACACGAAGGTTGAAGAGA 255
Db 389 CAGGCCCAACCAACAAAGCGGGCTCCACCAAGACCGCCACCAAGTAGAGGGCGA 448
QY 256 GGGAGAAGATACGATGCTGCCAGTGTGCGGTAGGATTTTCAATTAATCGAGAG 315
Db 449 GGCAGAAGGATCCGAATGCCGCCACGTCGCGGAGGATCTTTTCAGCTGACCGAGAG 508
QY 316 TTAGGTCACAAAATCCGACGCGGAACGATTCGGTGTGTTGGAGACGCTAGCGCGG 375
Db 509 CTCGGTCAATAATCCGACGCGGAACCATCCCGGTGGCTCCTCGAGACGCGCGCGC 568
QY 376 ATTATAGCGCCACCGGTACGGGAACGTTCCCGGCATCGCCATGTCGGTTAACGGAAC 435
Db 569 ATTATTCGACGACCGGCAACGCGACAGGTTCGGCCATCCGATGTTCCGCAATTGGAACG 628
QY 436 TTAATAATCCGACGAC 452
Db 629 TAAAGATTTCGACCAC 645

```

Search completed: February 2, 2004, 02:56:33
Job time : 2486.95 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

M nucleic - nucleic search, using sw model

un on: February 2, 2004, 01:11:01; Search time 4063.8 seconds
(without alignments)
10781.591 Million cell updates/sec

title: US-09-938-842A-1034

effect score: 1071

sequence: 1 atggcgacaattcagaagct.....cacggctcatcgaaaccactga 1071

scoring table: OLIGO NUC

Gapop 60.0, Gapext 60.0

searched: 2888711 seqs, 20454813386 residues

ord size: 0

total number of hits satisfying chosen parameters: 5777422

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: listing first 45 summaries

atabase: GenEmbl:

- 1: gb.ba.*
- 2: gb.hg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.vi.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.or.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB | ID | Description |
|------------|-------|---------------|--------|----|-----------|--------------------|
| 1 | 1071 | 100.0 | 1071 | 6 | AX506339 | Sequence |
| 2 | 1071 | 100.0 | 1071 | 8 | AF370606 | Arabidops |
| 3 | 1071 | 100.0 | 1102 | 8 | AY113981 | Arabidops |
| 4 | 1071 | 100.0 | 1660 | 8 | AY056214 | Arabidops |
| 5 | 1071 | 100.0 | 91854 | 8 | AC003680 | Arabidops |
| 6 | 25 | 2.3 | 134151 | 8 | AC006922 | Arabidops |
| 7 | 23 | 2.1 | 1206 | 8 | BT008493 | Arabidops |
| 8 | 23 | 2.1 | 1479 | 8 | AY081344 | Arabidops |
| 9 | 23 | 2.1 | 85961 | 8 | AC079131 | Arabidops |
| 10 | 23 | 2.1 | 103637 | 8 | AC079604 | Arabidops |
| 11 | 22 | 2.1 | 2134 | 3 | AF352244 | Arabidops |
| 12 | 22 | 2.1 | 149452 | 2 | AC124912 | Arabidops |
| 13 | 22 | 2.1 | 170737 | 2 | AC125510 | Arabidops |
| 14 | 21 | 2.0 | 158514 | 2 | AC051640 | Arabidops |
| 15 | 21 | 2.0 | 160796 | 9 | AL499604 | Human DNA |
| 16 | 21 | 2.0 | 166983 | 9 | AC002541 | Human DNA |
| 17 | 20 | 1.9 | 1099 | 10 | AF176350 | Rattus no |
| 18 | 20 | 1.9 | 1233 | 10 | AF218846 | Rattus no |
| 19 | 20 | 1.9 | 1305 | 10 | AF383875 | Mus muscu |
| 20 | 20 | 1.9 | 2709 | 10 | AF286721 | Rattus no |
| 21 | 20 | 1.9 | 2926 | 10 | AF286722 | Rattus no |
| 22 | 20 | 1.9 | 75493 | 2 | AC015877 | Homo sapi |
| 23 | 20 | 1.9 | 94810 | 6 | AX695407 | Sequence |
| 24 | 20 | 1.9 | 99007 | 2 | AC142506 | Medicago |
| 25 | 20 | 1.9 | 110000 | 2 | AC121458 | Continuation (2 of |
| 26 | 20 | 1.9 | 110000 | 2 | AC121458 | Continuation (3 of |
| 27 | 20 | 1.9 | 110000 | 2 | LMFLCHR36 | Continuation (22 o |
| 28 | 20 | 1.9 | 124347 | 9 | AC010072 | Continuation (22 o |
| 29 | 20 | 1.9 | 131701 | 9 | AC110583 | Homo sapi |
| 30 | 20 | 1.9 | 136064 | 10 | AL928892 | Homo sapi |
| 31 | 20 | 1.9 | 136150 | 8 | AP002485 | Mouse DNA |
| 32 | 20 | 1.9 | 136579 | 9 | HS998020 | Oryza sat |
| 33 | 20 | 1.9 | 141036 | 2 | AC023807 | Human DNA |
| 34 | 20 | 1.9 | 141227 | 2 | AC137836 | Medicago |
| 35 | 20 | 1.9 | 144820 | 2 | AC099556 | Trypanoso |
| 36 | 20 | 1.9 | 146301 | 2 | AC021414 | Homo sapi |
| 37 | 20 | 1.9 | 149042 | 2 | AC016410 | Homo sapi |
| 38 | 20 | 1.9 | 152990 | 2 | AC130432 | Homo sapi |
| 39 | 20 | 1.9 | 160359 | 9 | AC007262 | Homo sapi |
| 40 | 20 | 1.9 | 160846 | 9 | AC124306 | Homo sapi |
| 41 | 20 | 1.9 | 169337 | 2 | AC123755 | Mus muscu |
| 42 | 20 | 1.9 | 170867 | 9 | AC068295 | Homo sapi |
| 43 | 20 | 1.9 | 172159 | 2 | AP005492 | Oryza sat |
| 44 | 20 | 1.9 | 178461 | 9 | AL357075 | Human DNA |
| 45 | 20 | 1.9 | 179155 | 10 | AL772401 | Mouse DNA |

ALIGNMENTS

RESULT 1
AX506339
LOCUS 1071 bp DNA linear PAT 27-SEP-2002
DEFINITION Sequence 1034 from Patent WO0216655.
ACCESSION AX506339
VERSION AX506339.1 GI:23387576
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi
1
REFERENCE
AUTHORS Harper, J.F., Kreps, J., Wang, X. and Zhu, T.
TITLE Stress-regulated genes of plants, transgenic plants containing

same, and methods of use
Patent: WO 0216555-A 1034 28-FEB-2002;
The Scripps Research Institute (US) ; Syngenta Participations AG
(CH)

JOURNAL
FEATURES
source
Location/Qualifiers
1. .1071
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"

BASE COUNT 287 a 288 c 256 g 240 t
ORIGIN

Query Match 100.0%; Score 1071; DB 6; Length 1071;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1071; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCGCAATTCAGAGCTTGAAGAGTTGTCAGGCAAGATCAAACTCTAAGAGCGTT 60
DB 1 ATGGCGCAATTCAGAGCTTGAAGAGTTGTCAGGCAAGATCAAACTCTAAGAGCGTT 60
QY 61 GATCTAACCATCATCAACGGCTCGAAGAGCTCGAAGAGCTTCCCAAGTAAT 120
DB 61 GATCTAACCATCATCAACGGCTCGAAGAGCTCGAAGAGCTTCCCAAGTAAT 120
QY 121 CCCACAGTGTCTCGAGCCCAAGCGGAGCGGTGATGCCGTCTTCAATGCTTTA 180
DB 121 CCCACAGTGTCTCGAGCCCAAGCGGAGCGGTGATGCCGTCTTCAATGCTTTA 180
QY 181 GCTCCACCGTCTCGACAGACACCATTCAGAGAGCTTCGACTAAGACCGTCAAG 240
DB 181 GCTCCACCGTCTCGACAGACACCATTCAGAGAGCTTCGACTAAGACCGTCAAG 240
QY 241 AAGGTTGAAGAGAGGAGGAGGATACGGATCGCTGCCACGCTGCGGTAGGATTTT 300
DB 241 AAGGTTGAAGAGAGGAGGAGGATACGGATCGCTGCCACGCTGCGGTAGGATTTT 300
QY 301 CAATTAAGTGTAGGTTAGTTCACAAATCCGAGCGGAACGATTCGGTGTGTTGGAG 360
DB 301 CAATTAAGTGTAGGTTAGTTCACAAATCCGAGCGGAACGATTCGGTGTGTTGGAG 360
QY 361 AAGCTGAGCGCGGATTTATAGCGCCACCGGTACGGGAACGGTTCGCGCATCGCATG 420
DB 361 AAGCTGAGCGCGGATTTATAGCGCCACCGGTACGGGAACGGTTCGCGCATCGCATG 420
QY 421 TCGGTTAAGCAACCTTAAATCCGAGCGGAGCAAGCTGATTCGATGAGTGAA 480
DB 421 TCGGTTAAGCAACCTTAAATCCGAGCGGAGCAAGCTGATTCGATGAGTGAA 480
QY 481 AATCTGATGAAGAGAAACGTTAAACGACCTTCTAAGTGTGATATAGACATAGCGAC 540
DB 481 AATCTGATGAAGAGAAACGTTAAACGACCTTCTAAGTGTGATATAGACATAGCGAC 540
QY 541 GCGGTTTACGTTCTCGGTTTGTAGTTCGAATTCGACGAGCAACGATCCAACTCG 600
DB 541 GCGGTTTACGTTCTCGGTTTGTAGTTCGAATTCGACGAGCAACGATCCAACTCG 600
QY 601 CAAGCTCTGCATCATCATCTGAGTGTGCTGACCACTTCTCGCGCAAGATGATCGATG 660
DB 601 CAAGCTCTGCATCATCATCTGAGTGTGCTGACCACTTCTCGCGCAAGATGATCGATG 660
QY 661 TGGGCTATTCCATCAACGCAATGATTCGACGCGTGGAGCTTTCTTTGATTCACAA 720
DB 661 TGGGCTATTCCATCAACGCAATGATTCGACGCGTGGAGCTTTCTTTGATTCACAA 720
QY 721 ATCGTGTGCTCGGATCAGCTCAGTTATAGCTTTTCCCGCGCGCTGCTCGCG 780
DB 721 ATCGTGTGCTCGGATCAGCTCAGTTATAGCTTTTCCCGCGCGCTGCTCGCG 780
QY 781 TCGTCTTACGTCGCGGTGTTCAACAGGTTCCACGATGGCTAGACCACTCTTTTACAA 840
DB 781 TCGTCTTACGTCGCGGTGTTCAACAGGTTCCACGATGGCTAGACCACTCTTTTACAA 840
QY 841 GTTGTTCAGAGAGCGGCTTTGTATCGGTTTCAGAGTTAGCGGTTTCGAATTTATCAAGA 900

DB 841 GTTGTTCAGAGAGCGGCTTTGTATCGGTTTCAGAGTTAGCGGTTTCGAATTTATCAAGA 900
QY 901 GCGAGCTCGGTTATGGCTCCGAGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCA 960
DB 901 GCGAGCTCGGTTATGGCTCCGAGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCA 960
QY 961 ATTGCAACAAACGAGCGCACACGCTGAGAGACTTCTCCCTAGAGATATACGAGAAACAA 1020
DB 961 ATTGCAACAAACGAGCGCACACGCTGAGAGACTTCTCCCTAGAGATATACGAGAAACAA 1020
QY 1021 GAGCTTACAGTTCATGAGCGCACACGAGCGGTCACGAGCGGTCAGACCACTCA 1071
DB 1021 GAGCTTACAGTTCATGAGCGCACACGAGCGGTCACGAGCGGTCAGACCACTCA 1071

RESULT 2
AF370606
LOCUS
DEFINITION
Arabidopsis thaliana putative PCF2-like DNA binding protein
AF2945680; F17K2.21) mRNA, complete cds.
ACCESSION
AF370606
VERSION
AF370606.1 GI:13877694
KEYWORDS
FLI CDNA.
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
REFERENCE
1 (bases 1 to 1071)
Lan, B., Southwick, A., Karlin-Neumann, G., Nguyen, M., Miranda, M.,
Palm, C.J., Bowser, L., Jones, T., Banh, J., Chen, H., Cheuk, R.,
Chung, M.K., Kim, C., Lin, J., Liu, S.X., Pham, P.K., Sakano, H.,
Shinn, P., Yamada, K., Ecker, J., Theologis, A. and Davis, R.W.
Direct Submission
Submitted (17-APR-2001) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
COMMENT
e-mail for correspondence: arab@sequence.stanford.edu
This clone was isolated by arab-PCR.
FEATURES
source
1. .1071
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="II"
/clone="CS15"
/note="ecotype: Columbia"
gene
1. .1071
/gene="At2g45680; F17K2.21"
CDS
1. .1071
/codon_start=1
/product="putative PCF2-like DNA binding protein"
/protein_id="AAK4325.1"
/db_xref="GI:13877695"
/translation="MARIQKLEEVAGKOOTLRAVDLTINGVNVETSRPFPNPTVS
LEPKAFVPSFMSLAPSPSTGPPPLKRASTKDRHKVGRGRIRMPATCAARIFQL
TRIGHKSGETIRWLENAEPAITATGTVPALAMSVNGTKLPTITNADSDMG
NLAKTKRPSNSEYIDIDAVASGSLAPIATTTIOPQALASTVAQOILLPQMG
PMWAI PSNAMIPTVGAFELIPIQAGSNOPQILAPFAAASPSYVA/VQAQSTWARD
PPIQVPSGFGVSVDSGNSLRATSWAPSSSGVTTGSSSSSIATTTTILRDFSL
ETYEKELQGFNPTTARSNH"

BASE COUNT 287 a 288 c 256 g 240 t
ORIGIN

Query Match 100.0%; Score 1071; DB 8; Length 1071;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1071; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCGCAATTCAGAGCTTGAAGAGTTGTCAGGCAAGATCAAACTCTAAGAGCGTT 60
DB 1 ATGGCGCAATTCAGAGCTTGAAGAGTTGTCAGGCAAGATCAAACTCTAAGAGCGTT 60


```
2Y 61 GATCTAACCATCATCAACGGCTCAGAAAGCTCGAAACTTCAAGACCTTTCAGTAAT 120
2b 61 GATCTAACCATCATCAACGGCTCAGAAAGCTCGAAACTTCAAGACCTTTCAGTAAT 120
2Y 121 CCCACAGTGAAGTCTGAGCCGAGCGGAGCGGCTGATCGCTGCTTTCATGCTTTA 180
2b 121 CCCACAGTGAAGTCTGAGCCGAGCGGAGCGGCTGATCGCTGCTTTCATGCTTTA 180
2Y 181 GTCACACCTCTTCGACGAGCACCACCATTAAGAGAGCTTCGACTAAGACCGTCAACG 240
2b 181 GTCACACCTCTTCGACGAGCACCACCATTAAGAGAGCTTCGACTAAGACCGTCAACG 240
2Y 241 AAGGTTGAAGAGGAGGAGGAGGATACCGATGCTCTGCGACCTGTCGGCTAGGATTTT 300
2b 241 AAGGTTGAAGAGGAGGAGGAGGATACCGATGCTCTGCGACCTGTCGGCTAGGATTTT 300
2Y 301 CAATTAACGAGAGTGAAGTCAAAATCCGACGCGGAGAAACGATTTCGTTGTTGAG 360
2b 301 CAATTAACGAGAGTGAAGTCAAAATCCGACGCGGAGAAACGATTTCGTTGTTGAG 360
2Y 361 AACGCTGAGCGCGGATATAGCGCGGCTACGGAGCGGTTCCGCGCATCGCATG 420
2b 361 AACGCTGAGCGCGGATATAGCGCGGCTACGGAGCGGTTCCGCGCATCGCATG 420
2Y 421 TCGGTTAAGCGAACCTTAAATCCGACGCGGAGAAACGATTTCGTTGTTGAG 480
2b 421 TCGGTTAAGCGAACCTTAAATCCGACGCGGAGAAACGATTTCGTTGTTGAG 480
2Y 481 AATCTGATGAGAGAGAGTAAACGCTTCTACAGTGAATATAGATATAGATAGGAC 540
2b 481 AATCTGATGAGAGAGAGTAAACGCTTCTACAGTGAATATAGATATAGATAGGAC 540
2Y 541 GCGGTTTACGTTTCCGTTTACCTTCAATTCGACGAGCAGCAACGATCCACCTCCG 600
2b 541 GCGGTTTACGTTTCCGTTTACCTTCAATTCGACGAGCAGCAACGATCCACCTCCG 600
2Y 601 CAGCTCTGCGATCATCACTGTGCTCAGGAACTTCTGCGCGAGGAAATGATCCGATG 660
2b 601 CAGCTCTGCGATCATCACTGTGCTCAGGAACTTCTGCGCGAGGAAATGATCCGATG 660
2Y 661 TGGGCTATTCCATCAAGCGCAATGATTCGAGCGGCTGCGAGCTTCTTCTGATCCCAA 720
2b 661 TGGGCTATTCCATCAAGCGCAATGATTCGAGCGGCTGCGAGCTTCTTCTGATCCCAA 720
2Y 721 ATCGCTGCTCGCTCGAATCAGCTTATAGCTTTTCCGCGCGGCTGCTTCGCGG 780
2b 721 ATCGCTGCTCGCTCGAATCAGCTTATAGCTTTTCCGCGCGGCTGCTTCGCGG 780
2Y 781 TCGTCTTACGTCGCGCTGTTCAACAGGCTTCCAGTGTGCTAGACCACTCTCTTACAA 840
2b 781 TCGTCTTACGTCGCGCTGTTCAACAGGCTTCCAGTGTGCTAGACCACTCTCTTACAA 840
2Y 841 GTTGTTCACAGCAGCGGCTTTGATTCGCTTTTACAGCTTACGCTTTCGAAATTCAGA 900
2b 841 GTTGTTCACAGCAGCGGCTTTGATTCGCTTTTACAGCTTACGCTTTCGAAATTCAGA 900
2Y 901 GCGAGCTCGGTTATGCTTCCGAGCTCAGCTCAGCGGTAAACCGGTAGTTCATCGTCA 960
2b 901 GCGAGCTCGGTTATGCTTCCGAGCTCAGCTCAGCGGTAAACCGGTAGTTCATCGTCA 960
2Y 961 ATTGCAACAAACGACGACACGCTGAGAGCTTCTCCCTAGAGATATACGAGAAACAA 1020
2b 961 ATTGCAACAAACGACGACACGCTGAGAGCTTCTCCCTAGAGATATACGAGAAACAA 1020
2Y 1021 GAGCTTACCGATTCATGAGACCAACACGACGCTCATCGACCACTGA 1071
2b 1021 GAGCTTACCGATTCATGAGACCAACACGACGCTCATCGACCACTGA 1071
```

RESULT 3
Y113981
CCUS

AY113981 1102 bp mRNA linear PLN 18-SEP-2002

DEFINITION Arabidopsis thaliana putative PCF2 DNA binding protein (At2g45680)
mRNA, complete cds.
ACCESSION AY113981
VERSION AV113981.1 GI:21281084
KEYWORDS FLI CDNA
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Arabidopsis thaliana
Sukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE 1 (bases 1 to 1102)
Yamada, K., Ban, J., Chan, M.M., Chang, C.H., Chang, E., Dale, J.M.,
Deng, J.M., Goldsmith, A.D., Lee, J.M., Onodera, C.S., Quach, H.L.,
Tang, C.C., Toriumi, M., Wu, H.C., Yamamura, Y., Yu, G., Bowser, L.,
Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J.,
Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B.,
Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T.,
Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K.,
Davis, R.W., Ecker, J.R. and Theologis, A.
Arabidopsis Open Reading Frame (ORF) Clones
Unpublished
REFERENCE 2 (bases 1 to 1102)
Yamada, K., Ban, J., Chan, M.M., Chang, C.H., Chang, E., Dale, J.M.,
Deng, J.M., Goldsmith, A.D., Lee, J.M., Onodera, C.S., Quach, H.L.,
Tang, C.C., Toriumi, M., Wu, H.C., Yamamura, Y., Yu, G., Bowser, L.,
Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J.,
Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B.,
Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T.,
Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K.,
Davis, R.W., Ecker, J.R. and Theologis, A.
Direct Submission
Submitted (16-MAY-2002) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
The RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL cDNA: RIKEN
Arabidopsis Full-Length cDNA): Seki, M., Narusaka, M., Ishida, J.,
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
Hayashizaki, Y. and Shinozaki, K.
The Salk, Stanford, PGEC (SSP) Consortium members constructed and
sequenced the PUNI (ORF) clones using the RAFL cDNAs: Yamada, K.,
Ban, J., Chan, M.M., Chang, C.H., Chang, E., Dale, J.M., Deng, J.M.,
Goldsmith, A.D., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C.,
Toriumi, M., Wu, H.C., Yamamura, Y., Yu, G., Bowser, L., Chen, H.,
Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C., Lam, B., Lin, J.,
Miranda, M., Nguyen, M., Palm, C.J., Shinn, P., Southwick, A.,
Davis, R.W., Ecker, J.R. and Theologis, A.
Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally
to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP
/PGEC) contributed equally to this work as PIs.

Annotation is based on the January 2002 version of the Arabidopsis
genome submitted to GenBank.

FEATURES
Location/Qualifiers
1..1102
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="2"
/clone="U10678"
/note="this clone is in PUNI 51."
ecotype: Columbia
1..1102
/gene="At2g45680"
1..1071
/gene="At2g45680"
/codon_start=1
/evidence=experimental
/product="putative PCF2 DNA binding protein"
/protein_id="AA045029.1"
/db_xref="GI:21281085"
/translation="MATIQLEEYAGKQDTLRAVDLTIIINGVRNVTSRFPQVNPVTS"

Gene
CDS


```
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="2"
/clone="RAF107-08-P04 (R10678)"
/note="This clone is in a modified pBluescript vector
(FLC-1) as a BamHI/XhoI insert.
ecotype: Columbia"
1..1660
/gene="At2g45680"
1..102
/gene="At2g45680"
103..1173
/gene="At2g45680"
/codon_start=1
/evidence=experimental
/product="putative PCF2 DNA binding protein"
/protein_id="AA107063.1"
/db_xref="GI:15810351"
/translation="MATTOKLEEVAGKDOTLRVLDLTIINGVRNVETSRPPQVNPVTS
LEPKAPVMPSPMSLAPSPSTGPKLEASTKDHETKVEGRIRMPATCAARIFQL
TRELGHKSDGTISWLENAEPALIAATGTGTVEAIAMVNGTILKIPITINADSDMGE
NLMKKRRPNSSEVLEISDAVSASSGLAPIATITITTOPPOLASSVVAQQLFPQGY
EMWATPSNMIPTVGAFLIPQIAGPSNPOLLAPPAAPSPSYVAQVQASTMAP
PPLQVPSGVSFVSDVSGSLRATSVMAPSSSGVTGSSSIATITHTLDRFSL
EYKQELHQFMSITARRSNH"
1174..1660
/gene="At2g45680"
464 a 400 C 385 G 411 T
3' UTR
ASE COUNT 464 a 400 C 385 G 411 T
RIGIN
Query Match 100.0%; Score 1071; DB 8; Length 1660;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1071; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 1 ATGGCGCAATTCAGAACTTCAGAACTTCAGGCAAGATCAAACTCTAAGAGCGGTT 60
b 103 ATGGCGCAATTCAGAACTTCAGAACTTCAGGCAAGATCAAACTCTAAGAGCGGTT 162
Y 61 GATCTAACCATCATCAACGGCGTTCAGAACTTCAGAACTTCAGAACTTCAGAACT 120
b 163 GATCTAACCATCATCAACGGCGTTCAGAACTTCAGAACTTCAGAACTTCAGAACT 222
Y 121 CCCACAGTAGTCTCGACGCCAGCGGCGGTGATGCGGTGATGCGGTGATGCGGT 180
b 223 CCCACAGTAGTCTCGACGCCAGCGGCGGTGATGCGGTGATGCGGTGATGCGGT 282
Y 181 GCTCCACCGTCTTCACAGGACCCATTCAGAGAGGCTTCAGAGAGGCTTCAGAG 240
b 283 GCTCCACCGTCTTCACAGGACCCATTCAGAGAGGCTTCAGAGAGGCTTCAGAG 342
Y 241 AAGTTGAAGGAAGGAGGAGGATCGGATGCTGCGGCGGTGCGGCGGTGCGG 300
b 343 AAGTTGAAGGAAGGAGGAGGATCGGATGCTGCGGCGGTGCGGCGGTGCGG 402
Y 301 CAATTAACTCGAGAGTTAGTTCAGAACTTCAGAACTTCAGAACTTCAGAACT 360
b 403 CAATTAACTCGAGAGTTAGTTCAGAACTTCAGAACTTCAGAACTTCAGAACT 462
Y 361 AACGCTGAGCGCGGATTAATAGCGGCCAGCGGTACGGGACCGGTTCCCGCAT 420
b 463 AACGCTGAGCGCGGATTAATAGCGGCCAGCGGTACGGGACCGGTTCCCGCAT 522
Y 421 TCGGTTACGGAACCTTAAATATCCGACGACGACGACGACGACGACGACGAC 480
b 523 TCGGTTACGGAACCTTAAATATCCGACGACGACGACGACGACGACGACGAC 582
Y 481 AATCTGATGAGAGAAACGTAACGACCTTCTAAGAGGATATAGACATAGAGGAC 540
b 583 AATCTGATGAGAGAAACGTAACGACCTTCTAAGAGGATATAGACATAGAGGAC 642
Y 541 CGCGTTTCAGTTCTCGGTTTTCAGTTTCAGTTTCAGTTTCAGTTTCAGTTTC 600
```

```
Db 643 GCCTTTTCCCTTCCTCGGTTTAGCTCCAAATTCGACGACGACGATCCAACTCCG 702
Qy 601 CAAGCTCTGGCATCATCCACTGTGGCTCAGCAACTTTCGCCCAAGGAATGATCCGATG 660
Db 703 CAAGCTCTGGCATCATCCACTGTGGCTCAGCAACTTTCGCCCAAGGAATGATCCGATG 762
Qy 661 TGGGCTATTCCATCAACCAAGCAATGATTCGACGCTCGAGCTTCTTCTTGATTCACAA 720
Db 763 TGGGCTATTCCATCAACCAAGCAATGATTCGACGCTCGAGCTTCTTCTTGATTCACAA 822
Qy 721 ATCGCTGGTTCGCTCGAATCAGCTCAGTTTATTAGTTTTCGCCGCGCGCTGCTTCGCG 780
Db 823 ATCGCTGGTTCGCTCGAATCAGCTCAGTTTATTAGTTTTCGCCGCGCGCTGCTTCGCG 882
Qy 781 TCGTCTTACGTCGCGCTGCTTCAACAGGCTTCCACGATGGCTAGACACCTCTTTACAA 840
Db 883 TCGTCTTACGTCGCGCTGCTTCAACAGGCTTCCACGATGGCTAGACACCTCTTTACAA 942
Qy 841 GTTGTTCACAGCAGCGGCTTCTATCCGTTTCAGACGTTTACGCTTTCGATTCGATTCACAA 900
Db 943 GTTGTTCACAGCAGCGGCTTCTATCCGTTTCAGACGTTTACGCTTTCGATTCACAA 1002
Qy 901 GCGACGTGCTTATGCTCGGCTCAGCTCAAGCTCAGCGCTTAAACCAACCGGTGATTCATGCTCA 960
Db 1003 GCGACGTGCTTATGCTCGGCTCAGCTCAAGCTCAGCGCTTAAACCAACCGGTGATTCATGCTCA 1062
Qy 961 ATTGCAACCAACCAACGACGCTGAGAGACTTCTCCTAGAGATATACGAGAAACAA 1020
Db 1063 ATTGCAACCAACCAACGACGCTGAGAGACTTCTCCTAGAGATATACGAGAAACAA 1122
Qy 1021 GAGCTTCACAGCTTATGAGACCAACCAACGCTTATCGAACCACTGA 1071
Db 1123 GAGCTTCACAGCTTATGAGACCAACCAACGCTTATCGAACCACTGA 1173
```

RESULT 5
AC003680/c

LOCUS Arabidopsis thaliana chromosome 2 BAC F17K2 genomic sequence, complete sequence.
AC003680
VERSION AC003680.3 GI:20197048
KEYWORDS HTG.
SOURCE Arabidopsis thaliana (thale cress)

ORGANISM

Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Brassicales; Brassicaceae; Arabidopsi

REFERENCE

AUTHORS Lin, X., Kaul, S., Town, C.D., Benito, M.-I., Creasy, T.H., Haas, B.J.,
Wu, D., Maiti, R., Rong, C.M., Koo, H., Fujii, C.Y., Utterback, T.R.,
Barnstead, M.E., Bowman, C.L., White, O., Nierman, W.C. and Fraser, C.M.
Arabidopsis thaliana chromosome 2 BAC F17K2 genomic sequence
Unpublished
2 (bases 1 to 91854)

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BAC clone F17K2 is from Arabidopsis thaliana chromosome 2 and is
near the molecular marker(s) FLS.
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of several methods: Gene
prediction programs including Genscan+ (Chris Burge,

<http://CCR-081.mit.edu/GENSCAN.html>, GeneMarkHMM (Mark Borodovsky, <http://genemark.biology.gatech.edu/GeneMark/>), GlimmerA (a variant of GlimmerW, see Mihaela Perce, <http://www.tigr.org/softlab/glimmer.htm>), and GeneSplicer (Mihaela Perce and Steven Salzberg, contact mperte@tigr.org), searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding TRMs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Location/Qualifiers
1. 91854
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="2"
/map="C1C02E07"
/clones="F17K2"
complement(1..1281)
/notes="overlap with BAC clone F4123 (AC002387.1..1281)."
57..747
/gene="At2g45480"
/notes="synonym: F17K2.27"
join(<57..242,335..>747)
/gene="At2g45480"
join(57..242,335..747)
/gene="At2g45480"
/notes="Protein sequence is in conflict with the conceptual translation"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAFI8607.3"
/db_xref="GI:20197053"
/translation="DESKPSVQKFFPEVSDKCLEAAKFSNKNQDIIARSRWKNNV NGSLFGKIHSPVTLQBRGCFVLQGVETDNEPGRCRIDGKKRCKDVLGSGKYCD KHRHKKKHPVDITNSHNAFSLPVTETAVRVPCKDGDQKHVSVMGITLPR VDEKSTSCSTDTITDTRALRGEDDDEEYLSLSPGV"
complement(1002..2388)
/gene="At2g45490"
/notes="synonym: F17K2.2; contains a protein kinase domain profile (PDOC00100)"
complement(join<1002..1112,1208..1280,1366..1580,1655..1750,1886..2128,2260..>2388))
/gene="At2g45490"
complement(join(1002..1112,1208..1280,1366..1580,1655..1750,1886..2128,2260..2388))
/gene="At2g45490"
/codon_start=1
/product="putative protein kinase"
/protein_id="AAC06151.1"
/db_xref="GI:2979542"
/translation="MSGKSTESDAGNTEKQWLSADPEIGRLGKGFGRVYLAREAKS KYIVALKVFKQEIYKHQRLRREMEICTSLRHENILRLFGWHDNERIFLILEYA HGGLYGLVKQGHLEQQAQVIASLSALAYCHGKCVIHEDIKPENLLDHEGLK IADFGVSSNKRKTCMTGLDYLAPEMVENDHDIYAVDNTWLGILCVFLYGNPPE AESQKDTFRILKIDLSFPLTPNVSEAKNLISQLVKDPKRLSIEKIMQHPWIVRN ADPFGVCASIDI"
complement(3414..5980)
/gene="At2g45500"
/notes="synonym: F17K2.3; predicted by genscan"
complement(join<3414..3616,3639..3825,3915..4032,4215,4582..4712,4769..4864,4948..5188,5352..5374,5506..5585,5670..>5980))
/gene="At2g45500"

CDS

complement(join(3414..3616,3639..3825,3915..4032,4215,4582..4712,4769..4864,4948..5188,5352..5374,5506..5585,5670..5980))
/gene="At2g45500"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAC06152.1"
/db_xref="GI:2979543"
/translation="MSFLRGIIDFSFILNEESKDPSSVSSSTSSSMNGIDGPVPT NERTAYKLGYFDLAKBEIAKGVRAEWHLDHALLHYRNAORIMNEATSTPSPYSI SSEKVRYSREKTSNMCNOVSEBLOALGVMSNKETVAVPSASVSTSPSRKTL SOKTPVAGGVATPRNEKDAASPVPKESGNYDDKLVEINNTIIVDRSPSVKDDV AGLNCAKALBMLVLPKARDLFTGLRRPARVTSLLGLLFGPFGKRLKALAVAS ESQATFNVSASSLTSKWDISIMSTRSTSNEASRLKSEFLIQDGVTSNPDVLII IGANKQELDDAVLRILVRIYVLPDPSNRKLLFKTKLQCPHSLSDGDIKYKE TEGKLYLCKIKHFRISQVTDKRYGSDLOALCEEAAMPFIREIGANILTIQANKVLN FVSQINVEVCLSLKGLINRCVHTGKIST"
6429..8288
/gene="At2g45510"
/notes="synonym: F17K2.4"
join(<6429..6950,7026..7310,7386..7724,7806..8006,8100..>8288)
/gene="At2g45510"
join(6429..6950,7026..7310,7386..7724,7806..8006,8100..8288)
/gene="At2g45510"
/codon_start=1
/product="putative cytochrome P450"
/protein_id="AAC06153.1"
/db_xref="GI:2979544"
/translation="MEITGIAITVAITFIVLCFTIYLMIRIFTGSKRNDKRYAPVH ATVPDLFDSHLDYDYTEIARSKPTVRPLSPGQSBILTADPRNVHILKTRFDNYSK GHSSRENMDLLGHGIPAVDGEKWRQOKLSSFEFSTVLDRDPSCVFERNASKLVGF VSEFALSGKAFDAQDILLMRCLTDSIKVGFVGLKLDPSKEGQFMEAFDEGNVAT SSRFDLMLKLPFNIGSOKLKSATIDKFVYSILITTKKELAKEQNTVVREDIL SRFVSEKQDPMNNDKYLDIILNFMIAKDTTALLSNFLYMLCKNPLVQEKIYOE LPDGRVSKGDNITYIAYAMGRNTYIWDGAEFKPERLWKLQGLFQEPSPFKFISHA GPRICLCKDPAYRQMKIVSMALLHFFRFKWDENSKVYKRMILTHVDGLHLCAIPR TST"
complement(8412..9632)
/gene="At2g45520"
/notes="synonym: F17K2.5; supported by full length cDNA:
Ceres:8220"
complement(join<8412..8678,8768..8903,8987..9120,9289..9356,9459..>9632))
/gene="At2g45520"
complement(join(8553..8678,8768..8903,8987..9120,9289..9356,9459..9612))
/gene="At2g45520"
/codon_start=1
/product="expressed protein"
/protein_id="AAC06155.1"
/db_xref="GI:2979546"
/translation="MGCKKREKRYLAHGGPARLPPPPDRSQDDVPSTIRLMN YTSFDSRSTQVVEKKEKKAEEVVVATSDGSDSVVKKKKKKRKNQMTDLRFEN ELAIDRSKRKKKKYWEAKQKKGKGTEDTLRENFPEHQIRFGDVGQAPLKLA VVFKARSTLSAQERLQAIDYRSRKGWTFARPGVIPAVMQ"
9408..9427
/rpt_family="(CA)n"
complement(9578..9632)
/rpt_family="(GA)n"
9822..10798
/gene="At2g45530"
/notes="synonym: F17K2.6; supported by full length cDNA:
Ceres:35578"
9822..10798
/gene="At2g45530"
9948..10670
/gene="At2g45530"
/codon_start=1
/product="expressed protein"
/protein_id="AAC06154.1"

FEATURES

source

misc_feature

gene

mRNA

CDS

gene

mRNA

CDS

gene

mRNA

repeat_region

repeat_region

gene

mRNA

CDS

| | | | |
|--------------|-------|--|-----------------|
| Db | 74411 | ATCGTGGTCCGTGGAATCAGCTCAGTATTTAGCTTTTCCGCGCGCTGCTTCGCG | 74395 |
| Qy | 781 | TGCTCTTACGTCCGCGCTTTCAACAGAGGCTTCCAGATGGCTAGACCACTCTTTACAA | 840 |
| Db | 74351 | TGCTCTTACGTCCGCGCTTTCAACAGAGGCTTCCAGATGGCTAGACCACTCTTTACAA | 74292 |
| Qy | 841 | GTGTGTTCAAGCAGGCGCTTTGTATCCGTTTCAGAGTTAGCGGTTTCAATTTATCAAGA | 900 |
| Db | 74291 | GTGTGTTCAAGCAGGCGCTTTGTATCCGTTTCAGAGTTAGCGGTTTCAATTTATCAAGA | 74232 |
| Qy | 901 | GCGAGCTCGGTTATGGCTCCGAGCTCAAGCTCAGCGGTAAACAACCGGTAGTTTCATCGTCA | 960 |
| Db | 74231 | GCGAGCTCGGTTATGGCTCCGAGCTCAAGCTCAGCGGTAAACAACCGGTAGTTTCATCGTCA | 74172 |
| Qy | 961 | ATTGCAACAACAGCAGCGACACGCTGAGAGACTTCTCCCTAGAGATATACGAGAAACAA | 1020 |
| Db | 74171 | ATTGCAACAACAGCAGCGACACGCTGAGAGACTTCTCCCTAGAGATATACGAGAAACAA | 74112 |
| Qy | 1021 | GAGCTTCACCACTTCATGAGCACCACACAGCAGGTTCATCGAACCACTGA | 1071 |
| Db | 74111 | GAGCTTCACCACTTCATGAGCACCACACAGCAGGTTCATCGAACCACTGA | 74061 |
| RESULT 6 | | | |
| AC006922 | | 134151 bp | linear |
| LOCUS | | Arabidopsis thaliana chromosome 2 clone T118 map | PLN 27-FEB-2002 |
| DEFINITION | | sequence. | complete |
| ACCESSION | | AC006922 | |
| VERSION | | AC006922.7 | GI:20197941 |
| KEYWORDS | | HTG. | |
| SOURCE | | Arabidopsis thaliana (chale cress) | |
| ORGANISM | | Arabidopsis thaliana | |
| | | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; | |
| | | Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; | |
| | | rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis. | |
| REFERENCE | | 1 (bases 1 to 134151) | |
| AUTHORS | | Lin, X., Kaul, S., Shea, T. P., Fujii, C. Y., Shen, M., VanAken, S. E., | |
| | | Barnstead, M. E., Mason, T. M., Bowman, C. L., Renning, C. M., | |
| | | Benito, M. I., Carrera, A. J., Creasy, T. H., Buell, C. R., Town, C. D., | |
| | | Nierman, W. C., Fraser, C. M. and Venter, J. C. | |
| JOURNAL | | Unpublished | |
| REFERENCE | | 2 (bases 1 to 134151) | |
| AUTHORS | | Lin, X. | |
| TITLE | | Direct Submission | |
| JOURNAL | | Submitted (09-MAR-2000) The Institute for Genomic Research, 9712 | |
| | | Medical Center Dr. Rockville, MD 20850, USA | |
| REFERENCE | | 3 (bases 1 to 134151) | |
| AUTHORS | | Town, C. D. and Kaul, S. | |
| TITLE | | Direct Submission | |
| JOURNAL | | Submitted (27-FEB-2002) The Institute for Genomic Research, 9712 | |
| | | Medical Center Dr. Rockville, MD 20850, USA, cdtown@tigr.org | |
| COMMENT | | On Oct 18, 1997 this sequence version replaced gi:6598638. | |
| | | On Oct 8, 1997 this sequence version replaced gi:2443866. | |
| | | We have determined that YAC YUP8H12 is chimeric, and is comprised | |
| | | of two distinct genomic EcoRI fragments from chromosome 1. This | |
| | | submission contains the sequence from the EcoRI site at position 1 | |
| | | (right end) to position 181918 of our previous Phase II | |
| | | submission YUP8H12 accession number AC000098. This fragment maps | |
| | | at the bottom of the right arm of the chromosome between the ATHATPAS | |
| | | and m532 markers. | |
| FEATURES | | Location/Qualifiers | |
| source | | 1..134151 | |
| | | /organism="Arabidopsis thaliana" | |
| | | /mol_type="genomic DNA" | |
| | | /cultivar="Columbia" | |
| | | /db_xref="taxon:3702" | |
| | | /chromosome="2" | |
| | | /map="g6825" | |
| | | /clones="T118" | |
| misc feature | | 1..2183 | |

```

/repeat_region
/AC006282:90190..92376).
/complement(228..251)
/rpt_family="AT_rich"
1600..1690
/rpt_family="AT_rich"
3774..3801
/rpt_family="AT_rich"
4362..4397
/rpt_family="AT_rich"
5421..5442
/rpt_family="AT_rich"
5481..5527
/rpt_family="AT_rich"
6605..7737
/gene="At2g36830"
/note="synonym: T1J8.1; supported by full length cDNA:
Ceres:36633"
Join(6605..7052,7144..7737)
6640..6996
/gene="At2g36830"
/note="molecular marker ve017"
Join(6669..7052,7144..7515)
/gene="At2g36830"
/codon_start=1
/product="putative aquaporin (tonoplast intrinsic protein
gamma)"
/protein_id="AAD31569.1"
/db_xref="GI:4883600"
/translation="MPTRTAIGRPDEATPDALKALAEFTSLIFVAVSGSGMAF
NKLTEWATPGSVLAAVAHAFGLFVAVSGVANGISGGHVNPAVTFGAFIGNITLLR
GILVIAQLGSLVACILKFAFGGLAVPAFGLSAGVGLNPAFVFIWTFGLVTVY
ATAIDPVGSLGTIAPATIGFIVGANTLAGAFSGASMPAVAFGPAVSWTWNHW
WYAGLVGGGIGLAGLIEVFVINTTHEQLPTDY"
7847..7891
/rpt_family="AT_rich"
7952..7981
/rpt_family="(A)n"
/complement(10803..12597)
/gene="At2g36835"
/note="synonym: T1J8.24; supported by full length cDNA:
Ceres: 97088"
complement(join(<10803..11068,11697..11765,12150..12260,
12343..>12597))
/gene="At2g36835"
complement(join(11018..11068,11697..11765,12150..12260,
12343..12486))
/gene="At2g36835"
/codon_start=1
/product="Expressed protein"
/protein_id="AAM15323.1"
/db_xref="GI:20197944"
/translation="MAKGGATLAEDAPRWVSGRGVVPKISRPSVLISQNPETDIYAI
SVNKHFPVGGFAMAEVLESAGPECVPGVQTPLALLGVKVPVEVDLKEPVPKE
LKMVGFMNAVLMNKSFTDR"
complement(12752..14805)
/gene="At2g36840"
/note="synonym: T1J8.2; predicted by genefinder; supported
by full length cDNA: Ceres:108427"
complement(join(<12752..13150,13249..13533,13650..13971,
14054..14159,14237..14464,14576..>14805))
/gene="At2g36840"
complement(join(13013..13150,13249..13533,13650..13971,
14054..14159,14237..14464,14576..14729))
/gene="At2g36840"
/codon_start=1
/product="expressed protein"
/protein_id="AAD31570.1"
/db_xref="GI:4883601"
/translation="MGILSDVVIIISQSEKGDPSVITNCPTDKTGLCDLCRILLFF
GLNIVSGDSTGCKVLYWVIGKNTRWNLKRLVASFSGVIGRCLYSDS
ESQPKPLDLFLKLACSDRTGLLYDTEVLYKLEINIEKVKISTTPDGKVDLPFTV

```

Gene

mRNA

CDS

Gene

tRNA

repeat_region

Gene

mRNA

CDS

repeat_region

repeat_region

repeat_region

repeat_region

```

DTRELLGTVRREVEYELDAIGDSMISYDIELVPEITACSTSSSSVAETLPSDDVS
GEHSGSLTSSNVSIAVDSNSSLASHTLHITCQDHKGLLYIMRTFKDFOISYGRF
TIKGLCNCEIDLFIQSDCRKILDSKLNALITRLRAELOQLPRLVMMNRGPDTELLV
TNPVLSGKGRPOVPHDIALALKKIDTCIFSAEIGHRTVGTGDREVEVKVLNEEDSLP
IPRSKIESEVWKTLMGWE"
complement(15868..27297)
/gene="At2g36850"
/note="synonym: T1J8.3"
complement(join(<15868..16473,16628..16678,16797..16916,
16997..17099,17634..17773,17867..18035,18602..18764,
18879..18921,19083..19201,19620..19735,20017..20136,
20301..20424,20823..20988,21072..21147,21292..21426,
21519..21628,21711..21871,21920..>27297))
/gene="At2g36850"
complement(join(16114..16473,16628..16678,16797..16916,
16997..17099,17634..17773,17867..18035,18602..18764,
18879..18921,19083..19201,19620..19735,20017..20136,
20301..20424,20823..20988,21079..21147,21292..21426,
21519..21628,21711..21871,21920..>27297))
/gene="At2g36850"
/codon_start=1
/product="putative glucan synthase"
/protein_id="AAD31571.1"
/db_xref="GI:4883602"
/translation="MFTPTACFLSWPMKTESFLSKPKGRKVSFTFOVKRLHLILIT
VKDAANVPKNLEARRLESTFSLNPDNQAAPVAPWVPSPVETPYSTVLYSSSE
LRSNEEGISILFYLOKIFDEVENFLERGRSESTGDAIOASDALSIRFWVSFR
QVTLATVRGMVYTRRALMQSLFERGLGVDDASLTNMPRGFESSLEAQAADLKPT
VSCQIYGQOKQOKPEATDIGLLQRYEARLVAFTHSDVGVGGSGGKKEFYSK
LVKADIHQBEEIYSIKLPGDKLGEKPNQNHAIVFTGEATQITDMQNDYLSBA
IKMNLLEFPHGKHGIRRTILGVREHYTGRVMHYGHDPDFDIFPHIIRGGISKAS
RVINISDIYAGMFNSTLQGNITHEDVGLNQIALFEKGVAGGAGLSDRVRI
GOLFDFRMSYFTTGVFVQWMAFSQADRLSRVAKLSGNTALGAALNAOLFQIG
LFTAPVMVGFILLEGILKXAIPSFITNQFCVFFPFSGLTTHFTFGRITLHGAKV
YLLLIAMTVEDFDWVSWLMYKGVGVKGLSNWSEEOAHIOIOTLGRILETILS
LRPFMFGIYIKLDRTRKTSALGYSWVILVWVIFKGVASITFTALIVATAM
TDSIPDMFACVLGFIPTGHALLSLATKQVLRVLGLNETVREFGRIYDAAMGLIF
SPALSWFPFISTFQSRLLFNOAFSRGLGISILAGNRANVT"
complement(31341..31424)
/gene="At2g36860"
complement(31341..31424)
/gene="At2g36860"
/product="tRNA-Tyr"
complement(32969..33006)
/rpt_family="(TA)n"
complement(33824..35833)
/gene="At2g36870"
/note="synonym: T1J8.5; similar to cellulase (xyloglucan
endo transglycosylase) GI:311835 from (Tropaeolum majus);
supported by full length cDNA: Ceres: 7831"
complement(join(<33824..34414,35094..35311,35394..35494,
35596..>35833))
/gene="At2g36870"
complement(join(34048..34414,35094..35311,35394..35494,
35596..35809))
/gene="At2g36870"
/codon_start=1
/product="xyloglucan endotransglycosylase, putative"
/protein_id="AAD31572.1"
/db_xref="GI:4883603"
/translation="MGNLSILSLISFLHVLAVLGSSVNAVWPPSGVWPSKVSGLNLFY
KGFNLNGHQHQMDQNALITWLDRTSGSGFKSVFRSGYFGANLKQPGITAGVIT
SLYSNNEAHAPGHDEVDIEFLGTTPGKYTIQTNVYINGSGDGKLIIGREMKFLWFD
PTXDFHAILMSPREIILFVDDIPIRYPPKKSATFFPLRPMWLYGSDWASWATED
GKVDYKQPTAKYTNFKALGCTAYSSARCYPLSAPYRSGGLTROHQHARWVQT
HSMVYKCKYKEDHSLTPECWR"
complement(34562..34600)
/rpt_family="AT_rich"
complement(36785..36911)
/rpt_family="AT_rich"
37494..37687
/rpt_family="AT_rich"
complement(37725..37745)

```



```

repeat_region 37801 37845 /rpt family="AT_rich"
repeat_region 38111 38146 /rpt family="AT_rich"
Gene complement (40567..42581)

```

```

Query Match 2.3%; Score 25; DB 8; Length 134151;
Best Local Similarity 100.0%; Pred. No. 0.42; Indels 0; Gaps 0;
Matches 25; Conservative 0; Mismatches 0;
2Y 222 GACTAAGACCGTCACACGAGGTT 246
Db 101694 GACTAAGACCGTCACACGAGGTT 101718

```

```

RESULT 7
BT008493 Arabidopsis thaliana At1g58100 gene, complete cds. PLN 15-MAY-2003
LOCUS Arabidopsis thaliana (thale cress)
DEFINITION Arabidopsis thaliana
ACCESSION BT008493
VERSION BT008493.1 GI:30725659
KEYWORDS FLI CDNA
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1206)
Cheuk,R., Chen,H., Kim,C.J., Shinn,P., Bowser,L., Carninci,P.,
Dale,J.M., Hayashizaki,Y., Hsuan,V.W., Ishida,J., Jones,T.,
Kamuya,A., Karlin-Neumann,G., Kawai,J., Lam,B., Lin,J., Miranda,M.,
Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L.,
Sakurai,T., Satou,M., Seki,M., Southwick,A., Toriumi,M., Wong,C.,
Wu,H.C., Yamada,K., Yu,G., Yuan,S., Shinozaki,K., Davis,R.W.,
Theologis,A. and Ecker,J.R.
Arabidopsis ORF clones
Unpublished
2 (bases 1 to 1206)
Cheuk,R., Chen,H., Kim,C.J., Shinn,P., Bowser,L., Carninci,P.,
Dale,J.M., Hayashizaki,Y., Hsuan,V.W., Ishida,J., Jones,T.,
Kamuya,A., Karlin-Neumann,G., Kawai,J., Lam,B., Lin,J., Miranda,M.,
Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L.,
Sakurai,T., Satou,M., Seki,M., Southwick,A., Toriumi,M., Wong,C.,
Wu,H.C., Yamada,K., Yu,G., Yuan,S., Shinozaki,K., Davis,R.W.,
Theologis,A. and Ecker,J.R.
Direct Submission
Submitted (15-MAY-2003) Salk Institute Genomic Analysis Laboratory
(SiGNAL), Plant Biology Laboratory, The Salk Institute for
Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
USA
RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL CDNA : 'RIKEN
Arabidopsis Full-length cDNA') : Seki,M., Narusaka,M., Ishida,J.,
Satou,M., Kamuya,A., Sakurai,T., Carninci,P., Kawai,J.,
Hayashizaki,Y. and Shinozaki,K.

```

```

TITLE Arabidopsis ORF clones
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1206)
AUTHORS Cheuk,R., Chen,H., Kim,C.J., Shinn,P., Bowser,L., Carninci,P.,
Dale,J.M., Hayashizaki,Y., Hsuan,V.W., Ishida,J., Jones,T.,
Kamuya,A., Karlin-Neumann,G., Kawai,J., Lam,B., Lin,J., Miranda,M.,
Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L.,
Sakurai,T., Satou,M., Seki,M., Southwick,A., Toriumi,M., Wong,C.,
Wu,H.C., Yamada,K., Yu,G., Yuan,S., Shinozaki,K., Davis,R.W.,
Theologis,A. and Ecker,J.R.
Direct Submission
Submitted (15-MAY-2003) Salk Institute Genomic Analysis Laboratory
(SiGNAL), Plant Biology Laboratory, The Salk Institute for
Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
USA
RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL CDNA : 'RIKEN
Arabidopsis Full-length cDNA') : Seki,M., Narusaka,M., Ishida,J.,
Satou,M., Kamuya,A., Sakurai,T., Carninci,P., Kawai,J.,
Hayashizaki,Y. and Shinozaki,K.

```

```

TITLE Arabidopsis ORF clones
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1206)
AUTHORS Cheuk,R., Chen,H., Kim,C.J., Shinn,P., Bowser,L., Carninci,P.,
Dale,J.M., Hayashizaki,Y., Hsuan,V.W., Ishida,J., Jones,T.,
Kamuya,A., Karlin-Neumann,G., Kawai,J., Lam,B., Lin,J., Miranda,M.,
Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L.,
Sakurai,T., Satou,M., Seki,M., Southwick,A., Toriumi,M., Wong,C.,
Wu,H.C., Yamada,K., Yu,G., Yuan,S., Shinozaki,K., Davis,R.W.,
Theologis,A. and Ecker,J.R.
Direct Submission
Submitted (15-MAY-2003) Salk Institute Genomic Analysis Laboratory
(SiGNAL), Plant Biology Laboratory, The Salk Institute for
Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
USA
RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL CDNA : 'RIKEN
Arabidopsis Full-length cDNA') : Seki,M., Narusaka,M., Ishida,J.,
Satou,M., Kamuya,A., Sakurai,T., Carninci,P., Kawai,J.,
Hayashizaki,Y. and Shinozaki,K.

```

The Salk, Stanford, PGC (SSP) Consortium members constructed and sequenced the PUMI (ORF) clones using the RAFL cDNAs: Cheuk,R., Chen,H., Kim,C.J., Shinn,P., Bowser,L., Chan,M.M., Chang,C.M., Dale,J.M., Hsuan,V.W., Jones,T., Karlin-Neumann,G., Lam,B., Lee,J.M., Lin,J., Miranda,M., Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Southwick,A., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C., Yamada,K., Yu,G., Yuan,S., Davis,R.W., Theologis,A., and Ecker,J.R.

Cheuk,R. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk) contributed equally to this work as PIs.

```

FEATURES
Location/Qualifiers
1..1206
/molecule="mRNA"

```

```

/db xref="taxon:3702"
/chromosome="1"
/clone="U24662"
/note="This clone is in pUNI 51
ecotype: Columbia"
1..1206
/note="unknown protein"
/codon_start=1
/product="At1g58100"
/protein_id="AAP37852.1"
/db xref="GI:30725660"
/translational="MDLSIRNNNDTAAVATCGGARQLVDASLSIVPSTPESTL
ATTSSTATATTTKSTKDRHTKVDGRIRMPALCAARVQUTRELGHKSGETIEW
LQGAFAIVATGTGTIPANFTLSVSRSGSTLSAPPSKSVPLVYGLGTHHGYD
EQGGVFAAHTSPILGHQHQHONQNDPVTIPEGENFSKRYRVDLSKEN
DRAKONENKSESTSGPTAAEMWAVAPSSAGNFTWMLPVPPTAGNOMSSNN
NTAAGHRAPMPWPFVNSAGGAGGGAATHFMAGTGFSPMDQVSGSLQLGSLPAQ
PQTONLGSLMPDPSNLGLMLAALNSAYSRGNANAEQANNAVEHKEKQQSDHDDDS
RENSNSSE"
BASE COUNT 328 a 283 c 326 g 269 t
ORIGIN

```

```

Query Match 2.1%; Score 23; DB 8; Length 1206;
Best Local Similarity 100.0%; Pred. No. 6.2; Indels 0; Gaps 0;
Matches 23; Conservative 0; Mismatches 0;
QY 220 TCGACTAAGACCGTCACACGAA 242
Db 175 TCGACTAAGACCGTCACACGAA 197

```

```

RESULT 8
AY081344 Arabidopsis thaliana 1479 bp mRNA linear PLN 24-MAR-2002
LOCUS Arabidopsis thaliana auxin-induced basic helix-loop-helix
DEFINITION Transcription factor, putative (At1g58100) mRNA, complete cds.
ACCESSION AY081344.1 GI:19698994
VERSION AY081344
KEYWORDS FLI CDNA
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1479)
Nguyen,M., Karlin-Neumann,G., Southwick,A., Lam,B., Miranda,M.,
Palm,C.J., Bowser,L., Jones,T., Banh,J., Carninci,P., Chen,H.,
Cheuk,R., Chung,M.K., Hayashizaki,Y., Ishida,J., Kamuya,A.,
Kawai,J., Kim,C., Lin,J., Liu,S.X., Narusaka,M., Pham,P.K.,
Sakano,H., Sakurai,T., Satou,M., Seki,M., Shinn,P., Yamada,K.,
Shinozaki,K., Ecker,J., Theologis,A. and Davis,R.W.
Direct Submission
Submitted (26-FEB-2002) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
e-mail for correspondence: arab@sequence.stanford.edu

```

```

TITLE Arabidopsis thaliana auxin-induced basic helix-loop-helix
JOURNAL Transcription factor, putative (At1g58100) mRNA, complete cds.
REFERENCE 1 (bases 1 to 1479)
AUTHORS Nguyen,M., Karlin-Neumann,G., Southwick,A., Lam,B., Miranda,M.,
Palm,C.J., Bowser,L., Jones,T., Banh,J., Carninci,P., Chen,H.,
Cheuk,R., Chung,M.K., Hayashizaki,Y., Ishida,J., Kamuya,A.,
Kawai,J., Kim,C., Lin,J., Liu,S.X., Narusaka,M., Pham,P.K.,
Sakano,H., Sakurai,T., Satou,M., Seki,M., Shinn,P., Yamada,K.,
Shinozaki,K., Ecker,J., Theologis,A. and Davis,R.W.
Direct Submission
Submitted (26-FEB-2002) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
e-mail for correspondence: arab@sequence.stanford.edu

```

RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL CDNA : 'RIKEN Arabidopsis Full-length cDNA') : Seki,M., Narusaka,M., Ishida,J., Satou,M., Kamuya,A., Sakurai,T., Carninci,P., Kawai,J., Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PGC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Nguyen,M., Southwick,A., Karlin-Neumann,G., Lam,B., Miranda,M., Palm,C.J., Bowser,L., Jones,T., Banh,J., Chen,H., Cheuk,R., Chung,M.K., Kim,C., Lin,J., Liu,S.X., Pham,P.K., Sakano,H., Shinn,P., Yamada,K., Ecker,J., Theologis,A. and Davis,R.W.

Nguyen,M. (SSP/Stanford) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Davis,R.W. (SSP/Stanford) contributed equally to this work as PIs.

[illegible]

```
gene
/rpt family="AT rich"
complement(5299..6853)
/gene="T18124.3"
/notes="contains similarity to bZIP transcriptional
activator RSG GI:8777512 from (Nicotiana tabacum)"
complement(join(<5299..5357,5415..5651,5738..5864,
5968..6043,6135..>6853))
/gene="T18124.3"
complement(join(5299..5357,5415..5651,5738..5864,
5968..6043,6135..>6853))
/gene="T18124.3"
/codon start=1
/product="hypothetical protein"
/protein id="AAG50761.1"
/db_xref="GI:12321383"
/translation="MSSSSGSSVNLNVPKHALLPKIPPPSVASYSYIPTGLI
GRHGKLNKTHQRTSSSHLVEELPFWLDDLLNEQPSRPAKCHRSSSDSYA
YLDVATNATISLNDQSYRNTVLTSGVQELDRNKAQDAAPYSGASFQKRSRQ
RSLVATACPSLWLPFAENGKNGKLGALYMSQDAIVTSERKNYAEFPFSDPKMWS
EENNNSPTVYADNTKAKQQAQRGRVRKLOVISELRNVOTLQAGSKVAEILD
FLNQRLNLSMVENKALKKRLSIAOEKLIKOYSTVLKIVILMSSVYVPEQVLEKEI
GLRALYQQOQOTKPSASGRATSKDLDSPSSLSLNKDSNCRDSDSVSNVGFQDST
RSDVLPPEHRRF"
complement(8504..8535)
/rpt family="AT rich"
complement(8599..8627)
/rpt family="AT rich"
10149..11703
/gene="T18124.4"
/notes="Predicted by genemark hm"
join(<10149..10243,10536..>11703)
/gene="T18124.4"
join(10149..10243,10536..11703)
/gene="T18124.4"
/codon start=1
/product="hypothetical protein"
/protein id="AAG50763.1"
/db_xref="GI:12321385"
/translation="MEIQMRQMLMSNLKSSRSVLTGGNHTAASADTFLKASRE
VLSFMRTLLALLFTSLTSLKLYGTATAPSKSVESDLPPELLPILLDLKQGLF
KMGKALFLSGDDDEVTSSYSQVETDMLLSASQEMQSMVDPDSTFPAHSPH
IDSAEFIDRTKVGQIFVQLNLODLPENLLKPNVEIVVKSSEYVNTWTKYGE
OKQSLVAGRKLGITDEDAEKALKLEDVLLPFAARSKSRTYFKRTYRIPDLNG
DNLLDSKRVRFIDVNGKSSGSEWFEVPTRNQKFEYKLETVDNEMSLSEKX
KSTWLNKSESVYVNLKAAEVAEVEEMRSKSKMVDLFLCKPLGRLGRKMQS
KSGRAYWECALYGLRDEGVAVHQWNG"
complement(10335..10360)
/rpt family="GAGAA)n"
complement(12815..13163)
/gene="T18124.17"
/notes="pseudogene, putative heat shock transcription
factor HSF30"
/pseudo
complement(13870..13965)
/rpt family="AT rich"
complement(14034..14074)
/rpt family="AT rich"
complement(14182..14205)
/rpt family="AT rich"
complement(14901..18863)
/gene="T18124.5"
/notes="similar to copia type polyprotein GI:6850900 from
(Arabidopsis thaliana)"
complement(<14901..>18863)
/gene="T18124.5"
complement(14901..18863)
/gene="T18124.5"
/codon start=1
/product="copia-type polyprotein, putative"
/protein id="AAG50765.1"
/db_xref="GI:12321387"
/translation="MAGNNVFPQVPLTKSNYNWSLRKALGAHDVMEIVKGFIE
PENEGSLSQTKOGLDRSRDKKALCLLYQGLDEDTFEKVEATSAKEAWKLRITSY
```

```
Query Match 2.1%; Score 23; DB 8; Length 85961;
Best Local Similarity 100.0%; Pred No. 4.4;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Ct 220 TCAGTAAAGACCGTCACACGAA 242
Db 3190 TCAGTAAAGACCGTCACACGAA 3168
```

```
RESULT 10
AC079604/c
LOCUS
DEFINITION
Arabidopsis thaliana chromosome 1 BAC T15M6 genomic sequence,
complete sequence.
AC079604
VERSION
AC079604.5 GI:12321249
KEYWORDS
HTG
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi
1 (bases 1 to 103637)
Lin.X., Kaul.S., Town.C.D., Benito,M.-I., Creasy.T.H., Haas,B.J.,
Wu,D., Maiti.R., Ranning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,
Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
Arabidopsis thaliana chromosome 1 BAC T15M6 genomic sequence
Unpublished
2 (bases 1 to 103637)
Town,C.D. and Kaul,S.
Direct Submission
Submitted (04-SEP-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
3 (bases 1 to 103637)
Town,C.D. and Kaul,S.
Direct Submission
Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
On Jan 19, 2001 this sequence version replaced gi:12280852.
Address all correspondence to:at@tigr.org
BAC clone T15M6 is from Arabidopsis thaliana chromosome 1
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of several methods: Gene
prediction programs including Genscan+ (Chris Burge,
http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHM (Mark Borodovsky,
http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant
of GlimmerM, see Michaela Pertea,
http://www.tigr.org/softlab/glimmerm.htm/glimmerm.html, and
```



```

CDS
57...>2134
/clone="Dert11"
/codon_start=1
/product="paramyosin-like allergen"
/protein_id="AAK3511.1"
/db_xref="GI:13785807"
/translation="NNKRDSELAQRKLKLEDVHIESEYAHRLKQHAQIQEMQDQ
LQLOKAKNSDKPEKQFAVPELLAQLEATANKELTALKNVKLEYVHSLNKIE

```

Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: <http://www.nisc.nih.gov>
Contact: nisc_zoo@hgrl.nih.gov
----- Project Information
Center project name: cyu
Center clone name: 205M12

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair

data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 148556 bases at least Q40
Consensus quality: 148787 bases at least Q30
Consensus quality: 148973 bases at least Q20
Insert size: 150000; agarose-fp
Insert size: 149152; sum-of-contigs
Quality coverage: 11.35x in Q20 bases; agarose-fp
Quality coverage: 11.42x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

* 18748 18847: contig of 18747 bp in length
* 18848 18847: gap of unknown length
* 127485 127484: contig of 108637 bp in length
* 127585 147017: contig of unknown length
* 147018 147117: gap of unknown length
* 147118 149452: contig of 2335 bp in length.

FEATURES

source

1. 149452
/organism="Papio anubis"
/mol_type="genomic DNA"
/db_xref="taxon:9555"
/clones="RP41-205M12"
/clone_lib="RP41"

1. 44256

/note="Clone overlaps with GenBank Accession Number AC125510 clone RP41-382N23 (center project name cyt)"

1. 18747

/note="assembly_fragment
clone end:SP6
vector side:left"

18848-127484

/note="assembly_fragment"
127585..147017

/note="assembly_fragment"
147118..149452

/note="assembly_fragment
clone end:T7
vector side:right"

BASE COUNT 46580 a 31101 c 29868 g 41603 t 300 others

ORIGIN

Query Match 2.1%; Score 22; DB 2; Length 149452;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 473 TGGGTGAATCTGATCAAGAA 494
|||||

Db 35684 TGGGTGAATCTGATCAAGAA 35705
|||||

RESULT 13

AC125510

LOCUS

170737 bp DNA linear

HTG 20-JUL-2002

DEFINITION

Papio anubis clone RP41-382N23, WORKING DRAFT SEQUENCE, 4 ordered pieces.

ACCESSION

AC125510

VERSION

HTG; HTGS PHASE2; HTGS DRAFT.

KEYWORDS

Papio anubis (olive baboon)

SOURCE

Papio anubis

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae; Papio

REFERENCE

1 (bases 1 to 170737)

AUTHORS

Akhter N., Antonellis A., Ayele K., Beckstrom-Sternberg S.M., Benjamin B., Blakesley R.W., Bouffard G.G., Breen K., Brinkley C., Brooks S., Dietrich N.L., Granite S., Guan X., Gupta J., Haghighi P., Hansen N., Ho S.-L., Idol J.R., Karlins E., Laric P., Lee-Lin S.-Q., Legaspi R., Maduro Q.L., Maduro V.B., Marquies E.H., Masello C., Maskeri B., Mastrian S.D., McCloskey J.C., McDowell J., Paguirigan C., Pearson R., Portnoy M.E., Prasad A., Schueler M.G., Stantrijop S., Thomas J.W., Thomas J.W., Touchman J.W., Tsugeon C., Vogt J.L., Walker M.A., Wetherby K.D., Wiggins L., Young A., Zhang L.-H. and Green E.D.

TITLE

NIH Comparative Sequencing Initiative

JOURNAL

Unpublished

AUTHORS

2 (bases 1 to 170737)

JOURNAL

Direct Submission

REFERENCE

Submitted (28-JUN-2002) NIH Intramural Sequencing Center, 8717

AUTHORS

3 (bases 1 to 170737)

JOURNAL

Direct Submission

REFERENCE

Submitted (20-JUL-2002) NIH Intramural Sequencing Center, 8717

AUTHORS

Government Circle, Gaithersburg, MD 20877, USA

JOURNAL

Direct Submission

COMMENT

On Jul 20, 2002 this sequence version replaced gi:21622721.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.zoo@nih.gov
----- Project Information
Center project name: cyt
Center clone name: 382N23

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 169910 bases at least Q40
Consensus quality: 170146 bases at least Q30
Consensus quality: 170296 bases at least Q20
Insert size: 162000; agarose-fp
Insert size: 170437; sum-of-contigs
Quality coverage: 11.73x in Q20 bases; agarose-fp
Quality coverage: 11.15x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

* This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 89919: contig of 89919 bp in length
 * 90020: gap of unknown length
 * 146139: contig of 56119 bp in length
 * 146139: gap of unknown length
 * 146239: contig of 17872 bp in length
 * 164111: gap of unknown length
 * 164211: contig of 6527 bp in length.

FEATURES

source
 1. 170737
 /organism="Papio anubis"
 /mol_type="genomic DNA"
 /db_xref="taxon:9555"
 /clone="RP41-382N23"
 /clone_lib="RP41"
 misc_feature
 1. 89919
 /note="assembly_fragment
 clone end:SP6
 vector side:left"
 misc_feature
 1. 38014
 /note="clone overlaps with GenBank Accession Number
 AC12511 clone RP41-427L1 (center project name cys)"
 90020..146138
 /note="assembly_fragment"
 132730..170737
 /note="assembly_fragment"
 AC124912 clone RP41-205M12 (center project name cyu)"
 146239..164110
 /note="assembly_fragment"
 164211..170737
 /note="assembly_fragment"
 clone end:T7
 vector side:right"
 ASE COUNT 55152 a 35046 c 32826 g 47413 t 300 others
 RIGIN

Query Match 2.1%; Score 22; DB 2; Length 170737;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 473 TGGGTGAAATCTGATGAAGA 494
 |||||
 b 162194 TGGGTGAAATCTGATGAAGA 162215

RESULT 14
 C051640
 OCUS AC051640 158514 bp DNA linear HTG 24-AUG-2002
 EPIFFION Homo sapiens chromosome 9 clone RP11-161B4 map 9, WORKING DRAFT
 SEQUENCE, 25 unordered pieces.
 AC051640
 AC051640.4 GI:9966320
 HTG; HTGS PHASE1; HTGS_DRAFT.
 BYWORDS Homo sapiens (human)
 SOURCE Homo sapiens

ORGANISM
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 1 (bases 1 to 158514)

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 158514)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
 Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G.,
 Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,

Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
 Levine,R., Lieu,G., Liu,G., Locke,K., Macdonald,P., Marguis,N.,
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
 Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Ollivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (15-APR-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 158514)

REFERENCE

AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavsky,L.,
 Bouckgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
 Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
 Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
 FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
 Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
 Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRoque,K.,
 Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G.,
 Macdonald,P., Marguis,N., McCarthy,M., McEwan,P., McKernan,K.,
 McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,
 Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
 O'Donnell,P., O'Neill,D., Ollivar,T.M., Oliver,J., Peterson,K.,
 Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
 Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
 Soungez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
 Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
 Zimmer,A. and Zody,M.

Direct Submission

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Sep 4, 2000 this sequence version replaced gi:7960427.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIER

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L8703

Center clone name: 161_B_4

----- Summary Statistics

Sequencing vector: M13; M77815; 93% of reads

Sequencing vector: Plasmid; n/a; %-0.fff of reads

6.50344295923489Chemistry: Dye-terminator Big Dye; 100% of

reads

Assembly program: Phrap; version 0.960731

Consensus quality: 136255 bases at least Q40

Consensus quality: 147839 bases at least Q30

Consensus quality: 152536 bases at least Q20

Insert size: 159000; agarose-fp

Insert size: 156114; sum-of-contigs

Quality coverage: 3.4 in Q20 bases; agarose-fp

Quality cov.

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 25 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1783: contig of 1783 bp in length

| | |
|-------------------------------------|--------------------------------------|
| * 1784 | 1883: gap of 100 bp |
| * 1884 | 3231: contig of 1348 bp in length |
| * 3232 | 3331: gap of 100 bp |
| * 3332 | 4352: contig of 1021 bp in length |
| * 4353 | 4452: gap of 100 bp |
| * 4453 | 5702: contig of 1250 bp in length |
| * 5703 | 5802: gap of 100 bp |
| * 5803 | 8393: contig of 2591 bp in length |
| * 8394 | 8493: gap of 100 bp |
| * 8494 | 10272: contig of 1779 bp in length |
| * 10273 | 10372: gap of 100 bp |
| * 10373 | 12752: contig of 2380 bp in length |
| * 12753 | 12852: gap of 100 bp |
| * 12853 | 15516: contig of 2664 bp in length |
| * 15517 | 15616: gap of 100 bp |
| * 15617 | 19562: contig of 3946 bp in length |
| * 19563 | 19662: gap of 100 bp |
| * 19663 | 33383: contig of 13721 bp in length |
| * 33384 | 33483: gap of 100 bp |
| * 33484 | 39548: contig of 5065 bp in length |
| * 39549 | 39648: gap of 100 bp |
| * 39649 | 43137: contig of 4489 bp in length |
| * 43138 | 43237: gap of 100 bp |
| * 43238 | 48586: contig of 5349 bp in length |
| * 48587 | 48686: gap of 100 bp |
| * 48687 | 55083: contig of 6397 bp in length |
| * 55084 | 55183: gap of 100 bp |
| * 55184 | 61762: contig of 6579 bp in length |
| * 61763 | 61862: gap of 100 bp |
| * 61863 | 72589: contig of 10727 bp in length |
| * 72590 | 72689: gap of 100 bp |
| * 72690 | 80806: contig of 8117 bp in length |
| * 80807 | 80906: gap of 100 bp |
| * 80907 | 88472: contig of 7566 bp in length |
| * 88473 | 88572: gap of 100 bp |
| * 88573 | 98624: contig of 10052 bp in length |
| * 98625 | 98724: gap of 100 bp |
| * 98726 | 107382: contig of 8658 bp in length |
| * 107383 | 107482: gap of 100 bp |
| * 107483 | 117644: contig of 10162 bp in length |
| * 117645 | 117744: gap of 100 bp |
| * 117745 | 128355: contig of 10611 bp in length |
| * 128356 | 128455: gap of 100 bp |
| * 128456 | 140324: contig of 11869 bp in length |
| * 140325 | 140424: gap of 100 bp |
| * 140425 | 153983: contig of 13559 bp in length |
| * 153984 | 154083: gap of 100 bp |
| * 154084 | 158514: contig of 4431 bp in length. |
| FEATURES | |
| source | |
| 1. 158514 | |
| /organism="Homo sapiens" | |
| /mol_type="genomic DNA" | |
| /db_xref="taxon:9606" | |
| /chromosome="9" | |
| /map="9" | |
| /clone="RP11-161B4" | |
| /clone_lib="RPCL-11 Human Male BAC" | |
| 1. 1783 | |
| /note="assembly_fragment | |
| clone end:SP6 | |
| vector_side:left" | |
| misc_feature 1894..3231 | |
| misc_feature 3332..4352 | |
| misc_feature 4453..5702 | |
| misc_feature 5803..8393 | |
| misc_feature 8494..10272 | |
| misc_feature 10373..12752 | |
| misc_feature 13753..15516 | |
| misc_feature 15617..19562 | |
| misc_feature 19663..33383 | |
| misc_feature 33484..39548 | |
| misc_feature 39649..43137 | |
| misc_feature 43238..48586 | |
| misc_feature 48687..55083 | |
| misc_feature 55184..61762 | |
| misc_feature 61863..72589 | |
| misc_feature 72690..80806 | |
| misc_feature 80907..88472 | |
| misc_feature 88573..98624 | |
| misc_feature 98725..107382 | |
| misc_feature 107383..117644 | |
| misc_feature 117645..128355 | |
| misc_feature 128356..140324 | |
| misc_feature 140325..153983 | |
| misc_feature 153984..154083 | |
| misc_feature 154084..158514 | |

abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; SWI, SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr-9> RP11-23B15 is from the library RP11-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACe3.6

This sequence is the entire insert of clone RP11-23B15 The true right end of clone RP11-54606 is at 35092 in this sequence.

FEATURES

Source

Location/Qualifiers

1..160796

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="9"

/clone="RP11-23B15"

/clone_lib="RP11-11.1"

289..503

/note="MER20 repeat: matches 1..214 of consensus"

629..915

/note="AluJo repeat: matches 1..286 of consensus"

2346..2870

/note="MLT1G repeat: matches 14..536 of consensus"

2876..3130

/note="L1MCS5 repeat: matches 6976..7233 of consensus"

3182..3503

/note="L1MCS5 repeat: matches 7282..7595 of consensus"

3504..3816

/note="AluX repeat: matches 1..310 of consensus"

3819..4135

/note="AluX repeat: matches 2..310 of consensus"

4136..4540

/note="L1MCS5 repeat: matches 7595..7936 of consensus"

5076..5127

/note="MIR repeat: matches 79..130 of consensus"

5144..5373

/note="MIR repeat: matches 2..260 of consensus"

6766..6870

/note="MER86 repeat: matches 25..132 of consensus"

6895..7062

/note="L1P83 repeat: matches 5964..6141 of consensus"

7468..7858

/note="L2 repeat: matches 2293..2709 of consensus"

8086..8450

/note="MER47A repeat: matches 2..366 of consensus"

8607..8714

/note="MER81 repeat: matches 1..107 of consensus"

9115..9302

/note="L2 repeat: matches 1610..1828 of consensus"

9462..9713

/note="L2 repeat: matches 2168..2419 of consensus"

9714..10159

/note="MLT2CB repeat: matches 1..501 of consensus"

10160..12130

/note="L2 repeat: matches 200..2168 of consensus"

12131..12507

/note="THE1C repeat: matches 1..371 of consensus"

12508..12552

/note="L2 repeat: matches 56..200 of consensus"

13576..13692

/note="MER63 repeat: matches 14..131 of consensus"

13762..14168

/note="MLT2FA repeat: matches 1..400 of consensus"

14228..14920

/note="SVA repeat: matches 552..1259 of consensus"

14273..15366

/note="CpG island"

/evidence=not experimental

14921..15936

/note="SVA repeat: matches 9..954 of consensus"

15940..16092

/note="L1M4 repeat: matches 3221..3374 of consensus"

16151..16438

/note="AluX repeat: matches 1..289 of consensus"

16534..16691

/note="MIR repeat: matches 18..184 of consensus"

17093..17158

/note="33 copies 2 mer aa 69% conserved"

17377..17426

/note="25 copies 2 mer aa 98% conserved"

18077..18230

/note="Charlie4 repeat: matches 1795..1952 of consensus"

18294..18606

/note="AluSq repeat: matches 1..313 of consensus"

18610..18663

/note="27 copies 2 mer ag 72% conserved"

18666..18971

/note="L1MB8 repeat: matches 5875..6168 of consensus"

19056..19295

/note="L1MBC repeat: matches 1655..1888 of consensus"

19285..19596

/note="L1M4 repeat: matches 5433..5742 of consensus"

19597..19715

/note="MLT11 repeat: matches 19..137 of consensus"

19772..20057

/note="AluSp/q repeat: matches 212..297 of consensus"

20319..20473

/note="MIR repeat: matches 59..210 of consensus"

20925..21126

/note="L1MB2 repeat: matches 5953..6164 of consensus"

21155..21464

/note="AluX repeat: matches 1..308 of consensus"

21485..21679

/note="5 copies 39 mer 69% conserved"

21730..22258

/note="CpG island"

22197..22396

/evidence=not experimental

22409..22522

/note="MIR repeat: matches 47..240 of consensus"

22741..22808

/note="MIR repeat: matches 25..145 of consensus"

23245..23554

/note="34 copies 2 mer ca 72% conserved"

23663..23826

/note="AluX repeat: matches 1..308 of consensus"

24288..24416

/note="L2 repeat: matches 1441..1600 of consensus"

24467..24492

/note="MIR repeat: matches 22..145 of consensus"

24712..24763

/note="13 copies 2 mer ac 96% conserved"

24765..24893

/note="36 copies 2 mer tt 76% conserved"

25072..25317

/note="MIR repeat: matches 52..188 of consensus"

25948..25993

/note="MIR repeat: matches 1..250 of consensus"

26307..26466

/note="2 copies 23 mer 100% conserved"

26543..26701

/note="MLT1D repeat: matches 5..166 of consensus"

26726..26839

/note="MIR repeat: matches 25..183 of consensus"

27352..27501

/note="MLT11 repeat: matches 19..134 of consensus"

27509..27576

/note="MIR repeat: matches 20..184 of consensus"

27681..27925

/note="34 copies 2 mer gg 66% conserved"

repeat_region /note="L2 repeat: matches 2407. .2668 of consensus"
27938. .28038
repeat_region /note="MIR repeat: matches 100. .262 of consensus"
29269. .29580
repeat_region /note="AluSq repeat: matches 1. .313 of consensus"
29600. .29946
repeat_region /note="L1MB8 repeat: matches 5817. .6171 of consensus"
29953. .30009
repeat_region /note="L2 repeat: matches 2671. .2729 of consensus"
30708. .31108
repeat_region /note="L1MB4 repeat: matches 5781. .6183 of consensus"
31109. .31166
repeat_region /note="29 copies 2 mer at 75% conserved"
31131. .31208
repeat_region /note="2 copies 39 mer 83% conserved"
31165. .31390
repeat_region /note="3 copies 42 mer 79% conserved"
31172. .31283
repeat_region /note="56 copies 2 mer at 84% conserved"
31294. .32004
repeat_region /note="L1MB4 repeat: matches 5081. .5786 of consensus"
32005. .32312
repeat_region /note="Aluub repeat: matches 1. .308 of consensus"

Query Match

Best Local Similarity 2.0%; Score 21; DB 9; Length 160796;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 244 GTTGAAGGAGGAGGAGGAGG 264

DB 12691 GTTGAAGGAGGAGGAGG 12711

Search completed: February 2, 2004, 07:55:25
Job time : 4077.3 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

DN nucleic - nucleic search, using sw model

Run on: February 1, 2004, 21:59:06 ; Search time 313.739 Seconds
(without alignments)
9214.976 Million cell updates/sec

File: US-09-938-842A-1034

Perfect score: 1071

Sequence: 1 atgggacaattcagaagct.....caggtcatcgaccactga 1071

Scoring table: OLIGO NUC

Gapop_60.0, Gapext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size: 0

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: N Geneseq 19Jun03.*

```

1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| result No. | Score | Query Match | Length DB | ID | Description |
|------------|-------|-------------|-----------|----|--------------------|
| 1 | 1071 | 100.0 | 1071 | 24 | Arabidopsis thalia |
| 2 | 289 | 27.0 | 460 | 25 | Arabidopsis thalia |
| 3 | 286 | 26.7 | 453 | 24 | Arabidopsis thalia |
| 4 | 21 | 2.0 | 755 | 24 | Arabidopsis thalia |
| 5 | 21 | 2.0 | 755 | 24 | Arabidopsis thalia |
| 6 | 20 | 1.9 | 453 | 22 | Oligonucleotide fo |
| 7 | 19 | 1.8 | 416 | 21 | Human Maguin-2 hom |
| 8 | 19 | 1.8 | 499 | 21 | Zea mays DNA fragm |
| | | | | | Arabidopsis thalia |

| | | | | | | | |
|---|----|----|-----|-------|----|----------|--------------------|
| c | 9 | 19 | 1.8 | 600 | 24 | ABQ43702 | Oligonucleotide fo |
| c | 10 | 19 | 1.8 | 600 | 24 | ABQ43703 | Oligonucleotide fo |
| c | 11 | 19 | 1.8 | 908 | 24 | ABL99906 | Human secretory po |
| c | 12 | 19 | 1.8 | 1156 | 21 | AAC39595 | Arabidopsis thalia |
| c | 13 | 19 | 1.8 | 1893 | 21 | AAC43180 | Arabidopsis thalia |
| c | 14 | 19 | 1.8 | 2794 | 24 | ABL99921 | Human secretory po |
| c | 15 | 19 | 1.8 | 15987 | 23 | ABL44198 | Human secretory po |
| c | 16 | 19 | 1.8 | 34094 | 20 | ABZ30163 | Drosophila melanog |
| c | 17 | 18 | 1.7 | 277 | 24 | ABL75342 | Complete nucleotid |
| c | 18 | 18 | 1.7 | 424 | 22 | ABZ35423 | Corn tassal-derive |
| c | 19 | 18 | 1.7 | 432 | 25 | ABX55556 | Human ovarian PCR- |
| c | 20 | 18 | 1.7 | 475 | 24 | ABL93299 | Bovine EST associa |
| c | 21 | 18 | 1.7 | 605 | 23 | ABL17167 | Arabidopsis thalia |
| c | 22 | 18 | 1.7 | 626 | 24 | ABQ66162 | Drosophila melanog |
| c | 23 | 18 | 1.7 | 645 | 21 | AAC45826 | Arabidopsis thalia |
| c | 24 | 18 | 1.7 | 674 | 24 | ABQ33616 | Arabidopsis thalia |
| c | 25 | 18 | 1.7 | 674 | 24 | ABQ33617 | Oligonucleotide fo |
| c | 26 | 18 | 1.7 | 692 | 23 | AAS80632 | Oligonucleotide fo |
| c | 27 | 18 | 1.7 | 721 | 22 | AAI96937 | DNA encoding novel |
| c | 28 | 18 | 1.7 | 848 | 21 | AAC45001 | Human neuroblastom |
| c | 29 | 18 | 1.7 | 852 | 24 | ABQ44878 | Arabidopsis thalia |
| c | 30 | 18 | 1.7 | 852 | 24 | ABQ44879 | Oligonucleotide fo |
| c | 31 | 18 | 1.7 | 954 | 21 | AAC49097 | Oligonucleotide fo |
| c | 32 | 18 | 1.7 | 1012 | 21 | AAC34099 | Arabidopsis thalia |
| c | 33 | 18 | 1.7 | 1029 | 24 | ABQ45262 | Arabidopsis thalia |
| c | 34 | 18 | 1.7 | 1029 | 24 | ABQ45263 | Oligonucleotide fo |
| c | 35 | 18 | 1.7 | 1049 | 21 | AAC39576 | Oligonucleotide fo |
| c | 36 | 18 | 1.7 | 1120 | 21 | AAC41903 | Arabidopsis thalia |
| c | 37 | 18 | 1.7 | 1390 | 22 | AAF60982 | Arabidopsis thalia |
| c | 38 | 18 | 1.7 | 1399 | 24 | ABQ47386 | P. putida K12440-a |
| c | 39 | 18 | 1.7 | 1399 | 24 | ABQ47387 | Oligonucleotide fo |
| c | 40 | 18 | 1.7 | 1449 | 24 | ABQ13552 | Oligonucleotide fo |
| c | 41 | 18 | 1.7 | 1449 | 24 | ABQ13553 | Oligonucleotide fo |
| c | 42 | 18 | 1.7 | 1452 | 21 | AAC38957 | Oligonucleotide fo |
| c | 43 | 18 | 1.7 | 1733 | 22 | AAK71547 | Arabidopsis thalia |
| c | 44 | 18 | 1.7 | 1733 | 22 | AAK71548 | Human immune/haema |
| c | 45 | 18 | 1.7 | 2271 | 23 | ABL19369 | Drosophila melanog |

ALIGNMENTS

```

RESULT 1
ABZ13229
ID ABZ13229 standard; DNA; 1071 BP.
XX
AC ABZ13229;
XX
DT 21-JAN-2003 (first entry)
XX
DE Arabidopsis thaliana stress regulated gene SEQ ID NO 1034.
XX
KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO200216655-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26695.
XX
PR 24-AUG-2000; 2000US-227866P.
PR 26-JAN-2001; 2001US-264647P.
PR 22-JUN-2001; 2001US-300111P.
XX
(SCRI ) SCRIPPS RES INST.
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI Harper JF, Krepes J, Wang X, Zhu T;
XX
DR WPI; 2002-304127/34.
XX

```


PT Identifying a stress condition to which a plant cell has been exposed
PT and producing plants with increased tolerance to these abiotic stresses
XX
PS Claim 144; SEQ ID NO 1034; 577pp + Sequence Listing; English.
XX
CC The invention relates to identifying a stress condition to which a plant
CC cell has been exposed, comprising:
CC (a) contacting nucleic acid representative of expressed polynucleotides
CC in the plant cell with an array or probes representative of the plant
CC cell genome; and
CC (b) detecting a profile of expressed polynucleotides in the plant cell
CC characteristic of a stress response. The method is useful in the
CC production of transgenic plants, cells and seeds and in producing plants
CC with increased tolerance to abiotic stress. The present sequence is that
CC of an Arabidopsis thaliana stress regulated gene (AB212196-AB217574) used
CC in methods of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office.
XX
SQ Sequence 1071 BP; 287 A; 288 C; 256 G; 240 T; 0 other;

Query Match 100.0%; Score 1071; DB 24; Length 1071;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1071; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCGCAATTCAGAGCTTGAAGAGTTGCGAGCAAGATCAAACTCTAAGACGGTT 60
DB 1 ATGCGCAATTCAGAGCTTGAAGAGTTGCGAGCAAGATCAAACTCTAAGACGGTT 60
QY 61 GATCTAACCATCATCAACGGCGTCAGAAAGCTCGAACTTCAAGACCTTCCAAAGTAAT 120
DB 61 GATCTAACCATCATCAACGGCGTCAGAAAGCTCGAACTTCAAGACCTTCCAAAGTAAT 120
QY 121 CCCACAGTGTCTCGAGCCCAAGCGCGCGGTGATGCGCGTTCCTCAATGTCTTTA 180
DB 121 CCCACAGTGTCTCGAGCCCAAGCGCGCGGTGATGCGCGTTCCTCAATGTCTTTA 180
QY 181 GCTCCACCGTCTTCGACGAGCACCACCAATTGAAGAGAGCTTCGACTAAAGACCGTCAACG 240
DB 181 GCTCCACCGTCTTCGACGAGCACCACCAATTGAAGAGAGCTTCGACTAAAGACCGTCAACG 240
QY 241 AAGTTTGAAGAGAGGAGAGATACGATGCTGCGCAAGTTCGGTGTGTTGGAG 360
DB 241 AAGTTTGAAGAGAGGAGAGATACGATGCTGCGCAAGTTCGGTGTGTTGGAG 360
QY 301 CAATTAACCTGAGAGTGTAGTCAAAATCCGACGCAAGTTCGGTGTGTTGGAG 360
DB 301 CAATTAACCTGAGAGTGTAGTCAAAATCCGACGCAAGTTCGGTGTGTTGGAG 360
QY 361 AACGCTGAGCGCGGATATAGCGGCAAGTTCGGTGTGTTGGAG 420
DB 361 AACGCTGAGCGCGGATATAGCGGCAAGTTCGGTGTGTTGGAG 420
QY 421 TCGGTTTAAAGAGAGTAAATCCGACGAGCAGTTCGATTCGATGAGTGA 480
DB 421 TCGGTTTAAAGAGAGTAAATCCGACGAGCAGTTCGATTCGATGAGTGA 480
QY 481 AATCTGATGAAGAGAAACGTAACGACCTTCTAAGTGTGATATAGACATAGCGAC 540
DB 481 AATCTGATGAAGAGAAACGTAACGACCTTCTAAGTGTGATATAGACATAGCGAC 540
QY 541 GCGGTTTCACTTCCCGGTTTGTGCTCAATGCGACGACGATTCACCTCG 600
DB 541 GCGGTTTCACTTCCCGGTTTGTGCTCAATGCGACGACGATTCACCTCG 600
QY 601 CAAGCTCTGCATCATCACTGTGGTTCAGCAACTTCTGCGCAAGGAATGATCCGATG 660
DB 601 CAAGCTCTGCATCATCACTGTGGTTCAGCAACTTCTGCGCAAGGAATGATCCGATG 660
QY 661 TGGGCTATTCCATCAACGCAATGATTCGACGGTGGAGCTTCTTCTTGTATCCAAA 720

DB 661 TGGGCTATTCCATCAACGCAATGATTCGACGGTGGAGCTTCTTCTTGTATCCAAA 720
QY 721 ATCGTGTGTGCGTTCGAATCAGCCTCAGTTATAGCTTTTCCCGCGCGCTTCCCG 780
DB 721 ATCGTGTGTGCGTTCGAATCAGCCTCAGTTATAGCTTTTCCCGCGCGCTTCCCG 780
QY 781 TCGTCTTAGTTCGCGCTTTCGAAGAGCTTCCAGAGTTCAGCAGCTTCCTTTTCAA 840
DB 781 TCGTCTTAGTTCGCGCTTTCGAAGAGCTTTCGAAGAGCTTTCCTTTTCAA 840
QY 841 GTTGTTCCTCAAGCAGCGCTTTCGAAGAGCTTTCGAAGAGCTTTCCTTTTCAA 900
DB 841 GTTGTTCCTCAAGCAGCGCTTTCGAAGAGCTTTCGAAGAGCTTTCCTTTTCAA 900
QY 901 GCGAGTGTGCTTATGCTTCGAGTTCAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCT 960
DB 901 GCGAGTGTGCTTATGCTTCGAGTTCGAGTTCAGCTTCAAGCTTCAAGCTTCAAGCT 960
QY 961 ATTGCAACAAACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1020
DB 961 ATTGCAACAAACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1020
QY 1021 GAGTTTCAACAGTTCATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1071
DB 1021 GAGTTTCAACAGTTCATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1071
RESULT 2
ABX61156/c
ID ABX61156 standard; DNA; 460 BP.
XX
AC ABX61156;
XX
DT 26-FEB-2003 (first entry)
XX
DE Arabidopsis thaliana polynucleotide #502.
XX
KW Thale cress; gene; ds; genetic manipulation; plant; biosynthesis;
KW genetic modification; environmental stress; disease resistance;
KW fungicide; insecticide; stress tolerance.
XX
OS Arabidopsis thaliana.
XX
PN US2002142319-A1.
XX
PD 03-OCT-2002.
XX
PF 07-AUG-2001; 2001US-0924035.
XX
PR 13-AUG-1999; 99US-149784P.
XX
PR 11-AUG-2000; 2000US-0638258.
XX
PA (GORL//) GORLACH J.
PA (ANY//) AN Y.
PA (HML//) HAMILTON C M.
PA (PRIC//) PRICE J L.
PA (HARG//) HARGISS T R.
PA (YUY//) YU Y.
PA (RAME//) RAMEAKA J G.
PA (PAGE//) PAGE A.
PA (MATH//) MATHW A V.
PA (LEDF//) LEDFORD B L.
PA (WOES//) WOESSNER J P.
PA (HAAS//) HAAS W D.
PA (GARC//) GARCIA C A.
XX
PI Gorlach J, An Y, Hamilton CM, Price JL, Hargiss TR, Yu Y,
PI Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;
PI Garcia CA,
XX
DR WPI; 2003-102509/09.
XX
PT Novel Arabidopsis thaliana nucleic acid useful for constructing a

transgenic plant with enhanced disease resistance and enhanced traits of interest, as probes, and in diagnosis and screening purposes -

Claim 1; Page 158; 277pp; English.

The invention relates to Arabidopsis thaliana nucleic acid sequences. The DNA sequences and the polypeptides they encode are useful for identifying homologous or related genes, for producing compositions that modulate the expression or function of the polypeptides, for mapping associated regions of the protein, in diagnosis, for studying associated physiological pathways, for genetic manipulation of cells, preferably plant cells, in screening assays of various plant strains to determine the strains that are capable of withstanding a particular disease or environmental stress, for enhancing or inhibiting production of biosynthetic products in plants and to create genetically modified and transgenic organisms, such as plant cells and plants. Transgenic plants are useful for introducing or improving disease resistance and stress tolerance in plants, screening biologically active agents, such as fungicides and insecticides, and for identifying factors involved in biosynthetic pathways of nutritional, commercial or medicinal value. Sequences ABX0655-ABX61554 represent Arabidopsis thaliana polynucleotides of the invention.

Sequence 460 BP; 113 A; 103 C; 134 G; 109 T; 1 other;

Query Match 27.0%; Score 289; DB 25; Length 460;

Best Local Similarity 99.6%; Pred. No. 1.4e-131;

Matches 459; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

600 GCAAGCTTGGCATCATCTGCTGGCTCAGCAATTTCTGGCGAAGAAATGTTATCCGAT 659

460 GCAAGCTTGGCATCATCTGCTGGCTCAGCAATTTCTGGCGAAGAAATGTTATCCGAT 401

660 GTGGGCTATTCCATCAAGCAATCATTCGCGAGCTTCCTTCTTGATTCACCA 719

400 GTGGGCTATTCCATCAAGCAATCATTCGCGAGCTTCCTTCTTGATTCACCA 341

720 AATCGCTGGTCGTCGATCAATCAAGCTTCAGTTATTAGCTTTTCCGCGCGCTGCTCGCC 779

340 AATCGCTGGTCGTCGATCAATCAAGCTTCAGTTATTAGCTTTTCCGCGCGCTGCTCGCC 282

780 GTGCTTACGTCGCGCTGTTCAACAGCTTCACAGATGCTAGACCACTCTTTTACA 839

281 GTGCTTACGTCGCGCTGTTCAACAGCTTCACAGATGCTAGACCACTCTTTTACA 222

840 AGTTGTTTCAAGCAGCGCTTGTATCGCTTTCAGACGTTAGCGGTTTCAATTTATCAAG 899

221 AGTTGTTTCAAGCAGCGCTTGTATCGCTTTCAGACGTTAGCGGTTTCAATTTATCAAG 162

900 AGCGAGCTGGTTATGGCTCCGAGCTCAAGCTCAGGCGTAACAACCGGTATTCATCGTC 959

161 AGCGAGCTGGTTATGGCTCCGAGCTCAAGCTCAGGCGTAACAACCGGTATTCATCGTC 102

960 AATTCCAAACAACAGCGCAGCTCAGAGATTTCTCCCTAGATATACGGAACA 1019

101 AATTCCAAACAACAGCGCAGCTCAGAGATTTCTCCCTAGATATACGGAACA 42

1020 AGAGTTTCAAGTTATGAGCAACCAACAGCAGCTGTCAT 1060

41 AGAGTTTCAAGTTATGAGCAACCAACAGCAGCTGTCAT 1

BL93850/c

D ABU93850 standard; cDNA; 453 BP.

X ABU93850;

X 10-JUN-2002 (first entry)

X Arabidopsis thaliana nucleic acid sequence Ref:2027615 SEQ ID NO:615.

X Arabidopsis thaliana; insecticide; fungicide; plant; mapping; diagnosis;

genetic modification; gene; ss.

Arabidopsis thaliana.

US2002023280-A1.

21-FEB-2002.

26-JAN-2001; 2001US-0770444.

27-JAN-2000; 2000US-178502P.

(GORLACH J.

(ANYI// AN Y.

(HAMI// HAMILTON C M.

(PRIC// PRICE J L.

(RAIN// RAINES T M.

(YUY// YU Y.

(RAME// RAMEAKA J G.

(PAGE// PAGE A.

(MATH// MATHW A V.

(LEDF// LEDFORD B L.

(WOES// WOESSNER J P.

(HAAS// HAAS W D.

(GARC// GARCIA C A.

(KRICK// KRICKER M.

(SLAT// SLATER T.

(DAVI// DAVIS K R.

(ALLE// ALLEN K.

(HOFF// HOFFMAN N.

(HURB// HURBAN P.

Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y,

Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD,

Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N,

Hurban P;

WPI; 2002-367486/31.

New Arabidopsis thaliana nucleic acid, for identifying homologous

genes, producing compositions that modulate the expression or function

of its encoded protein, and mapping functional regions of a protein -

Claim 1; SEQ ID 615; 44pp; English.

The present invention describes an Arabidopsis thaliana nucleic acid (I)

comprising a sequence capable of hybridising under stringent conditions

to a sequence (S1) selected from any one of the 999 sequences given in

ABL93236 to ABL94234. (I) have insecticide and fungicide activities, and

they can be used as protein expression modulators. (I) can be used in

identifying homologous or related genes, in producing compositions that

modulate the expression or function of their encoded proteins, mapping

functional regions of the proteins, and in studying associated

physiological pathways. (I) can also be used: (1) for the genetic

manipulation of cells, particularly plant cells; (2) in screening assays

of various plant strains to determine the strains that are best capable

of withstanding a particular disease or environmental stress; (3) for

enhancing or inhibiting production of a biosynthetic product in a plant;

(4) as probes in mapping and in diagnosis, in genetic modification and

for screening purposes, to generate additional copies of the nucleic

acids, to generate ribozymes or antisense oligonucleotides, and as

single-stranded DNA probes or as triple-strand forming oligonucleotides;

and (5) for generating genetically modified transgenic organisms.

Note: The sequence data for this patent did not form part of the printed

specification, but was obtained in electronic format directly from the

USPTO web site.

Sequence 453 BP; 111 A; 100 C; 132 G; 107 T; 3 other;

Query Match 26.7%; Score 286; DB 24; Length 453;

Best Local Similarity 100.0%; Pred. No. 4.3e-130;

Matches 286; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|------|----|---|------|
| 775 | QY | TCGCGCTGCTTACGTCGGCGCTGTTTCAACAGCGTTCCAGATGGCTAGACCACTCCT | 834 |
| 286 | Db | TCGCGCTGCTTACGTCGGCGCTGTTTCAACAGCGTTCCAGATGGCTAGACCACTCCT | 227 |
| 835 | QY | TTACAAGTTGTTTCCAGCAGCGGCTTTGTAATCCGTTTTACAGCGTTAGCGGTTTCGAATTTA | 894 |
| 226 | Db | TTACAAGTTGTTTCCAGCAGCGGCTTTGTAATCCGTTTTACAGCGTTAGCGGTTTCGAATTTA | 157 |
| 895 | QY | TCAAGAGCGACGTCGGTTATGGCTCCGAGTCAAGCTCAGCGTTAAACAACGGTAGTTCA | 954 |
| 166 | Db | TCAAGAGCGACGTCGGTTATGGCTCCGAGTCAAGCTCAGCGTTAAACAACGGTAGTTCA | 107 |
| 955 | QY | TCGTCAATTGCAACAACAACGACGCAACGCTCGAGAGCTTCTCCCTAGAGATATACGAG | 1014 |
| 106 | Db | TCGTCAATTGCAACAACAACGACGCAACGCTCGAGAGCTTCTCCCTAGAGATATACGAG | 47 |
| 1015 | QY | AAACAAGAGTTTCAACGAGTTATGAGCAACAACAGCAGCGTCAAT | 1060 |
| 46 | Db | AAACAAGAGTTTCAACGAGTTATGAGCAACAACAGCAGCGTCAAT | 1 |

| | |
|-------------|---|
| RESULT 4 | |
| ABQ47102/c | |
| ID | ABQ47102 standard; DNA; 755. BP. |
| XX | |
| XX | |
| ABQ47102; | |
| AC | |
| XX | |
| XX | |
| 12-JUL-2002 | (first entry) |
| XX | |
| DE | Oligonucleotide for detecting cytosine methylation SEQ ID NO 33693. |
| XX | |
| Human; | cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis; |
| KW | drug; side effect; cancer; central nervous system; cardiovascular; |
| KW | gastrointestinal; respiratory system; single nucleotide polymorphism; |
| KW | SNP; cell differentiation; ds. |

polymorphisms (SNP's); and (ii) for differentiation of cell or tissue types and for investigating cell differentiation. The method allows the methylation status of many C residues to be determined simultaneously. ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the method for determining the degree of cytosine methylation described in the disclosure of the invention.

Sequence 755 BP; 125 A; 78 C; 282 G; 270 T; 0 other;

Query Match 2.0%; Score 21; DB 24; Length 755;
Best Local Similarity 100.0%; Pred.No.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

438 AAAAAATCCGACGACGACGAA 458
88 AAAAAATCCGACGACGACGAA 68

ABQ47103 standard; DNA; 755 BP.

ABQ47103;

12-JUL-2002 (first entry)

Oligonucleotide for detecting cytosine methylation SEQ ID NO 33694.

Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis; drug; side effect; cancer; central nervous system; cardiovascular; gastrointestinal; respiratory system; single nucleotide polymorphism; SNP; cell differentiation; ds.

Homo sapiens.

WO200218632-A2.

07-MAR-2002.

01-SEP-2001; 2001WO-BP10074.

01-SEP-2000; 2000DE-1043826.

05-SEP-2000; 2000DE-1044543.

(EPIG-) EPIGENOMICS AG.

Olek A, Piepenbrock C, Berlin K, Guetig D;

WPI; 2002-371829/40.

Determining the degree of cytosine methylation in genomic DNA, useful for diagnosis and prognosis, comprises selective hybridization of amplicons from chemically treated DNA -

Claim 12; 56pp + Sequence Listing; 56pp; German.

This invention describes a novel method for determining the degree of methylation of a particular cytosine in a motif 5'-CpG-3', present in a genomic sample of DNA. The sample is treated chemically to convert cytosine (C) but not methylated C to uracil, then part of the genomic DNA that contains the target C is amplified to form a labeled amplicon. The amplicon is hybridised to two classes, each with at least one member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the degree of hybridisation to both classes is determined from the label on the amplicon. From the ratio of labels hybridised to the two classes of oligomers, the degree of methylation is calculated. The method is used: (i) for diagnosis and/or prognosis of side effects of therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders of the central nervous, cardiovascular, gastrointestinal and respiratory systems etc., particularly by detecting mutations or single nucleotide polymorphisms (SNP's); and (ii) for differentiation of cell or tissue types and for investigating cell differentiation. The method allows the methylation status of many C residues to be determined simultaneously.

ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the method for determining the degree of cytosine methylation described in the disclosure of the invention.

Sequence 755 BP; 270 A; 282 C; 78 G; 125 T; 0 other;
Query Match 2.0%; Score 21; DB 24; Length 755;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

438 AAAAAATCCGACGACGACGAA 458

668 AAAAAATCCGACGACGACGAA 688

RESULT 6

ABA08516/C

ID ABA08516 standard; cDNA; 453 BP.

ABA08516;

11-JAN-2002 (first entry)

Human Maguin-2 homologue-encoding cDNA, SEQ ID NO:292.

Human; cytokine; cell proliferation; cell differentiation; growth factor; haematopoiesis regulation; tissue growth; immunomodulator; activin; inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis; proliferation; metastasis; cancer; tumour; haematopoietic disorder; myeloid cell disorder; lymphoid cell disorder; asthma; arthritis; chronic inflammatory condition; proliferative retinopathy; atherosclerosis; coronary heart disease; arterial ischaemia; bone disorder; osteoporosis; vascular growth disorder; tissue regeneration; wound healing; infection; immune disorder; cell culture; drug screening; gene therapy; antiinflammatory; antiasthmatic; antiarthritis; haemostatic; antiarteriosclerotic; cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial; antifungal; vulnery; antiulcer; ss.

Homo sapiens.

WO200157188-A2.

09-AUG-2001.

05-FEB-2001; 2001WO-US03800.

03-FEB-2000; 2000US-0495914.

27-APR-2000; 2000US-0560875.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Drmanac RT;

WPI; 2001-457740/49.

P-PSDB; ABB11272.

Human proteins and DNA encoding sequences useful for preventing, treating or ameliorating a medical condition in a mammalian subject e.g. arthritis and cancer -

Claim 1; Page 438; 1963pp; English.

Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and sequences ABA08225-ABA09574 represent nucleic acids encoding them. The invention also relates to vectors and recombinant host cells comprising a nucleotide of the invention, methods of producing the novel polypeptides, antibodies against the polypeptides, methods of detecting the nucleotides or polypeptides in a sample, and methods of identifying compounds which bind to polypeptides of the invention. Although novel, many of the polypeptides of the invention have homology to known proteins, thereby giving an insight into their probable biological activities, and hence potential therapeutic applications. The polypeptides of the invention may

have various activities, including cytokine, cell proliferation or cell differentiation activities; stem cell growth factor activity; haematopoiesis regulatory activity; tissue growth activity; immunomodulatory activity; activin- or inhibin-related activities; chemotactic or chemokinetic activities; haemostatic, thrombotic or thrombolytic activities; receptor or ligand activities; or may be involved in oncogenesis, cancer cell proliferation or metastasis. Depending on their biological activities, polypeptides and nucleotides of the invention are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene therapy. Such conditions include cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell disorders), chronic inflammatory conditions (e.g., asthma or arthritis), proliferative retinopathy, atherosclerosis, coronary heart disease, arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal vascular growth. Polypeptides involved with tissue regeneration and repair (or nucleic acids encoding them) may be used to promote wound healing (e.g., of burns, incisions and ulcers), while those with immunomodulatory activities may be used in the treatment of viral, bacterial and fungal infections in addition to immune disorders. Polypeptides with growth factor activity may be used in cell cultures to promote cell growth. For example, such polypeptides may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease or accidental damage. The polypeptides and nucleotides may also be used in the diagnosis of the above conditions, and in drug screening techniques. The present sequence represents a cDNA encoding a novel human polypeptide of the invention.

Sequence 453 BP; 109 A; 101 C; 119 G; 124 T; 0 other;

Query Match 1.9%; Score 20; DB 22; Length 453;

Best Local Similarity 100.0%; Pred. No. 16;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 618 CACTGGGCTCAGCAACTTC 637

Db 239 CACTGGGCTCAGCAACTTC 220

RESULT 7

AAC41435

ID AAC41435 standard; DNA; 416 BP.

AAC41435;

17-OCT-2000 (first entry)

Zea mays DNA fragment SEQ ID NO: 31870.

Hybridisation assay; genetic mapping; gene expression control; protein identification; signal transduction pathway; metabolic; pathway; promoter; termination sequence; corn; ss.

Zea mays subsp. mays.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

25-FEB-1999; 99US-0121825.

05-MAR-1999; 99US-0123180.

09-MAR-1999; 99US-0123548.

23-MAR-1999; 99US-0125788.

25-MAR-1999; 99US-0126264.

29-MAR-1999; 99US-0126785.

01-APR-1999; 99US-0127462.

08-APR-1999; 99US-0128234.

16-APR-1999; 99US-0128714.

19-APR-1999; 99US-0129845.

21-APR-1999; 99US-0130077.

21-APR-1999; 99US-0130449.


```
14-OCT-1999; 99US-0159637.
14-OCT-1999; 99US-0159638.
18-OCT-1999; 99US-0159584.
21-OCT-1999; 99US-0160741.
21-OCT-1999; 99US-0160767.
21-OCT-1999; 99US-0160768.
21-OCT-1999; 99US-0160770.
21-OCT-1999; 99US-0160814.
21-OCT-1999; 99US-0160815.
22-OCT-1999; 99US-0160980.
22-OCT-1999; 99US-0160981.
22-OCT-1999; 99US-0160989.
25-OCT-1999; 99US-0161404.
25-OCT-1999; 99US-0161405.
26-OCT-1999; 99US-0161359.
26-OCT-1999; 99US-0161360.
26-OCT-1999; 99US-0161361.
28-OCT-1999; 99US-0161920.
28-OCT-1999; 99US-0161992.
28-OCT-1999; 99US-0161993.
29-OCT-1999; 99US-0162142.

Query Match      i.8; Score 19; DB 21; Length 416;
Best Local Similarity 100.0%; Pred.No. 50;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 761 CCGCGCGCGCTGCTCGCC 779
b 57 CCGCGCGCGCTGCTCGCC 75

RESULT 8
AC36816
D AAC36816 standard; DNA; 499 BP.
X C
X AAC36816;
C X
T 17-OCT-2000 (first entry)
T X
X Arabidopsis thaliana DNA fragment SEQ ID NO: 15176.
E X
E Hybridisation assay; genetic mapping; gene expression control;
W Protein identification; signal transduction pathway;
W Metabolic pathway; promoter; termination sequence; ss.
X S Arabidopsis thaliana.
X S
X EP1033405-A2.
N X
D 06-SEP-2000.
X X
F 25-FEB-2000; 2000EP-0301439.
X F
R 25-FEB-1999; 99US-0121825.
R 05-MAR-1999; 99US-0123180.
R 09-MAR-1999; 99US-0123548.
R 23-MAR-1999; 99US-0125788.
R 25-MAR-1999; 99US-0126264.
R 29-MAR-1999; 99US-0126785.
R 01-APR-1999; 99US-0127462.
R 06-APR-1999; 99US-0128234.
R 08-APR-1999; 99US-0128714.
R 16-APR-1999; 99US-0129845.
R 19-APR-1999; 99US-0130077.
R 21-APR-1999; 99US-0130449.
R 23-APR-1999; 99US-0130510.
R 23-APR-1999; 99US-0130891.
R 30-APR-1999; 99US-0131449.
R 30-APR-1999; 99US-0132048.
R 30-APR-1999; 99US-0132407.
R 04-MAY-1999; 99US-0132484.
R 05-MAY-1999; 99US-0132485.

06-MAY-1999; 99US-0132486.
06-MAY-1999; 99US-0132487.
07-MAY-1999; 99US-0132863.
11-MAY-1999; 99US-0134256.
14-MAY-1999; 99US-0134218.
14-MAY-1999; 99US-0134219.
14-MAY-1999; 99US-0134221.
14-MAY-1999; 99US-0134370.
18-MAY-1999; 99US-0134768.
18-MAY-1999; 99US-0134941.
20-MAY-1999; 99US-0135124.
21-MAY-1999; 99US-0135353.
24-MAY-1999; 99US-0135629.
25-MAY-1999; 99US-0136021.
27-MAY-1999; 99US-0136392.
28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
```


is used: (i) for diagnosis and/or prognosis of side effects of therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders of the central nervous, cardiovascular, gastrointestinal and respiratory systems etc., particularly by detecting mutations or single nucleotide polymorphisms (SNP's); and (ii) for differentiation of cell or tissue types and for investigating cell differentiation. The method allows the methylation status of many C residues to be determined simultaneously. ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the method for determining the degree of cytosine methylation described in the disclosure of the invention.

Sequence 600 BP; 136 A; 53 C; 216 G; 195 T; 0 other;

Query Match 1.8%; Score 19; DB 24; Length 600;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

441 AATCCGACGACGACGAC 459
238 AATCCGACGACGACGAC 220

RESULT 10

ABQ43703
ABQ43703 standard; DNA; 600 BP.

ABQ43703;

12-JUL-2002 (first entry)

Oligonucleotide for detecting cytosine methylation SEQ ID NO 30294.

Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis; drug; side effect; cancer; central nervous system; cardiovascular; gastrointestinal; respiratory system; single nucleotide polymorphism; SNP; cell differentiation; ds.

Homo sapiens.

WO200218632-A2.

07-MAR-2002.

01-SEP-2001; 2001WO-EP10074.

01-SEP-2000; 2000DE-1043826.

05-SEP-2000; 2000DE-1044543.

(EPIG-) EPIGENOMICS AG.

Olek A, Piepenbrock C, Berlin K, Guetig D;

WPI; 2002-371829/40.

Determining the degree of cytosine methylation in genomic DNA, useful for diagnosis and prognosis, comprises selective hybridization of amplicons from chemically treated DNA

Claim 12; 56pp + Sequence Listing; 56pp; German.

This invention describes a novel method for determining the degree of methylation of a particular cytosine in a motif 5'-CpG-3', present in a genomic sample of DNA. The sample is treated chemically to convert cytosine (C) but not methylated C, to uracil, then part of the genomic DNA that contains the target C is amplified to form a labeled amplicon. The amplicon is hybridised to two classes, each with at least one member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the degree of hybridisation to both classes is determined from the label on the amplicon. From the ratio of labels hybridised to the two classes of oligomers, the degree of methylation is calculated. The method is used: (i) for diagnosis and/or prognosis of side effects of therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders of the central nervous, cardiovascular, gastrointestinal and respiratory

systems etc., particularly by detecting mutations or single nucleotide polymorphisms (SNP's); and (ii) for differentiation of cell or tissue types and for investigating cell differentiation. The method allows the methylation status of many C residues to be determined simultaneously. ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the method for determining the degree of cytosine methylation described in the disclosure of the invention.

Sequence 600 BP; 195 A; 216 C; 53 G; 136 T; 0 other;

Query Match 1.8%; Score 19; DB 24; Length 600;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

441 AATCCGACGACGACGAC 459
363 AATCCGACGACGACGAC 381

RESULT 11

ABL99906/C

ID ABL99906 standard; cDNA; 908 BP.

ABL99906;

03-OCT-2002 (first entry)

Human secretory polynucleotide (sptm) 161.

Human; ss; gene; secretory protein; secretory polynucleotides; SPTM; SPTM-related disease; somatic gene therapy; germline gene therapy; severe combined immunodeficiency; intracellular parasite protection; fungal parasite; protozoan parasite; cell proliferative disorder; cancer; immune disorder; AIDS; neurological disorder; Parkinson's disease; motor neuron disorder; demyelinating disease; multiple sclerosis; meningitis; abscess; prion diseases; cerebral palsy; neuroskeletal disorder; peripheral nervous system disorder; dermatomyositis; polymyositis; myopathy; myasthenia gravis; mental disorder; Tourette's syndrome.

Homo sapiens.

WO200220756-A2.

14-MAR-2002.

30-AUG-2001; 2001WO-US27297.

05-SEP-2000; 2000US-229747P.

05-SEP-2000; 2000US-229748P.

05-SEP-2000; 2000US-229749P.

05-SEP-2000; 2000US-229750P.

05-SEP-2000; 2000US-229751P.

05-SEP-2000; 2000US-230016P.

05-SEP-2000; 2000US-230583P.

06-SEP-2000; 2000US-230505P.

06-SEP-2000; 2000US-230514P.

06-SEP-2000; 2000US-230515P.

06-SEP-2000; 2000US-230517P.

06-SEP-2000; 2000US-230518P.

06-SEP-2000; 2000US-230519P.

06-SEP-2000; 2000US-230595P.

06-SEP-2000; 2000US-230596P.

06-SEP-2000; 2000US-230597P.

06-SEP-2000; 2000US-230599P.

06-SEP-2000; 2000US-230610P.

06-SEP-2000; 2000US-230864P.

06-SEP-2000; 2000US-230865P.

06-SEP-2000; 2000US-230988P.

06-SEP-2000; 2000US-230989P.

06-SEP-2000; 2000US-230990P.

07-SEP-2000; 2000US-230896P.

07-SEP-2000; 2000US-230897P.

PR 07-SEP-2000; 2000US-230951P.
PR 07-SEP-2000; 2000US-231163P.
PR 07-SEP-2000; 2000US-231832P.
XX PA (INCY-) INCYTE GENOMICS INC.
XX
XX Stuart J, Lincoln SE, Altus CM, Dufour GE, Chalup MG, Hillman JL;
PI Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;
PI Moniyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;
PI Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A;
PI Marwaha R, Chen AJ, Chang SC, Au AP, Iman RR;
XX
XX WPI; 2002-315658/35.
DR P-PSDB; ABB97909.
XX
XX Polynucleotide sequences encoding human secretory proteins useful for
PT Gene therapy of e.g. genetic deficiency disorders, cancers, and
PT diseases caused by intracellular parasites -
XX
XX Claim 1; Page 349; 585pp; English.
XX
XX The invention comprises the amino acid and coding sequences of human
secretory (SPTM) proteins. The SPTM DNA and amino acid sequences are
useful for treating a disease or condition associated with the expression
of functional SPTM. The SPTM DNA sequences are useful for somatic or
germline gene therapy to correct a genetic deficiency (e.g. severe
combined immunodeficiency). The SPTM DNA sequences are also useful in
providing protection against intracellular parasites (e.g. fungal
parasites and protozoan parasites). The SPTM DNA and protein sequences
are also useful for diagnosing cell proliferative disorders, cancer,
immune disorders (e.g. AIDS), neurological disorders (e.g. Parkinson's
disease), motor neuron disorders, demyelinating diseases (e.g. multiple
sclerosis), meningitis, abscesses, prion diseases, cerebral palsy,
neurological disorders, peripheral nervous system disorders,
dermatomyositis and polymyositis, myopathy, myasthenia gravis, and mental
disorders (e.g. Tourette's syndrome). cDNA sequences ABL9746 - ABL99929
represent human secretory polynucleotides of the invention.
XX
XX Sequence 908 BP; 225 A; 214 C; 184 G; 285 T; 0 other;
SQ

Query Match Similarity 1.8%; Score 19; DB 24; Length 908;
Best Local Similarity 100.0%; Pred.No. 48;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 988 AGAGACTTCCTCCCTAGAGA 1006
|||||
Db 170 AGAGACTTCCTCCCTAGAGA 152

RESULT 12
AAC39595
ID AAC39595 standard; DNA; 1156 BP.
XX AC
XX AAC39595;
XX
XX 17-OCT-2000 (first entry)
DT
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 25210.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
PD
XX 25-FEB-2000; 2000EP-0301439.
FF
XX 25-FEB-1999; 99US-0121825.
XX
XX 05-MAR-1999; 99US-0123180.
PR

09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135134.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.

PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 28-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.8%; Score 19; DB 21; Length 1156;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 CTAAGACCGTCACACGAA 242
|||||
DB 287 CTAAGACCGTCACACGAA 305
|||||

RESULT 13
AAC43180
ID AAC43180 standard; DNA; 1893 BP.
XX
AC AAC43180;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 38322.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
DD 06-SEP-2000.
XX
FF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125786.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131149.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 18-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.

PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 23-OCT-1999; 99US-0161404.
 PR 23-OCT-1999; 99US-0161405.
 PR 23-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 1.8%; Score 19; DB 21; Length 1893;

Best Local Similarity 100.0%; Pred. No. 45;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 553 TCCTCGGTTTACGCCAA 571
 b 859 TCCTCGGTTTACGCCAA 877

RESULT 14

D ABL99921/c
 D ABL99921 standard; cDNA; 2794 BP.

X ABL99921;

X 03-OCT-2002 (first entry)

X Human secretory polynucleotide (sptm) 176.

Human; ss; gene; secretory protein; secretory polynucleotides; SPTM;
 SPTM-related disease; somatic gene therapy; germline gene therapy;
 severe combined immunodeficiency; intracellular parasite protection;
 immune disorder; AIDS; neurological disorder; Parkinson's disease;
 fungal parasite; protozoan parasite; cell proliferative disorder; cancer;
 motor neuron disorder; demyelinating disease; multiple sclerosis;
 meningitis; abscess; prion diseases; cerebral palsy;
 neuroskeletal disorder; peripheral nervous system disorder;
 dermatomyositis; polynucleotides; myopathy; myasthenia gravis;
 mental disorder; Tourette's syndrome.

X Homo sapiens.

X WO200220756-A2.

X 14-MAR-2002.

X 30-AUG-2001; 2001WO-US27297.

X 05-SEP-2000; 2000US-229747P.

X 05-SEP-2000; 2000US-229748P.

X 05-SEP-2000; 2000US-229749P.

X 05-SEP-2000; 2000US-229750P.

X 05-SEP-2000; 2000US-229751P.

X 05-SEP-2000; 2000US-230016P.

X 05-SEP-2000; 2000US-230583P.

X 06-SEP-2000; 2000US-230505P.

X 06-SEP-2000; 2000US-230514P.

PR 06-SEP-2000; 2000US-230515P.
 PR 06-SEP-2000; 2000US-230517P.
 PR 06-SEP-2000; 2000US-230518P.
 PR 06-SEP-2000; 2000US-230519P.
 PR 06-SEP-2000; 2000US-230595P.
 PR 06-SEP-2000; 2000US-230596P.
 PR 06-SEP-2000; 2000US-230597P.
 PR 06-SEP-2000; 2000US-230599P.
 PR 06-SEP-2000; 2000US-230610P.
 PR 06-SEP-2000; 2000US-230864P.
 PR 06-SEP-2000; 2000US-230865P.
 PR 06-SEP-2000; 2000US-230988P.
 PR 06-SEP-2000; 2000US-230989P.
 PR 06-SEP-2000; 2000US-230990P.
 PR 07-SEP-2000; 2000US-230896P.
 PR 07-SEP-2000; 2000US-230897P.
 PR 07-SEP-2000; 2000US-230951P.
 PR 07-SEP-2000; 2000US-231163P.
 PR 07-SEP-2000; 2000US-231832P.

(INCY-) INCYTE GENOMICS INC.

Stuart J, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;
 Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;
 Momiyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;
 Gerstin BH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A;
 Marwaha R, Chen AL, Chang SC, Au AP, Inman RR;

WPI; 2002-315658/35.

P-PSDB; ABB97925.

Polynucleotide sequences encoding human secretory proteins useful for
 gene therapy of e.g. genetic deficiency disorders, cancers, and
 diseases caused by intracellular parasites -

Claim 1; Page 360-361; 585pp; English.

The invention comprises the amino acid and coding sequences of human
 secretory (SPTM) proteins. The SPTM DNA and amino acid sequences are
 useful for treating a disease or condition associated with the expression
 of functional SPTM. The SPTM DNA sequences are useful for somatic or
 germline gene therapy to correct a genetic deficiency (e.g. severe
 combined immunodeficiency). The SPTM DNA sequences are also useful in
 providing protection against intracellular parasites (e.g. fungal
 parasites and protozoan parasites). The SPTM DNA and protein sequences
 are also useful for diagnosing cell proliferative disorders, cancer,
 immune disorders (e.g. AIDS), neurological disorders (e.g. Parkinson's
 disease), motor neuron disorders, demyelinating diseases (e.g. multiple
 sclerosis), meningitis, abscesses, prion diseases, cerebral palsy,
 neuroskeletal disorders, peripheral nervous system disorders,
 dermatomyositis and polynucleotides, myopathy, myasthenia gravis,
 disorders (e.g. Tourette's syndrome). cDNA sequences ABL99746 - ABL99929
 represent human secretory polynucleotides of the invention.

Sequence 2794 BP; 704 A; 656 C; 549 G; 885 T; 0 other;

Query Match 1.8%; Score 19; DB 24; Length 2794;

Best Local Similarity 100.0%; Pred. No. 44;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 988 AGAGACTTCTCCTAGAGA 1006

Db 1901 AGAGACTTCTCCTAGAGA 1883

RESULT 15

ABLI4198

ID ABLI4198 standard; cDNA; 15987 BP.

XX ABLI4198;

AC ABLI4198;

XX 26-MAR-2002 (first entry)

DT 26-MAR-2002 (first entry)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

4 nucleic - nucleic search, using sw model
on on: February 2, 2004, 00:53:06 ; Search time 2477.95 Seconds
(without alignments)
10504.671 Million cell updates/sec

File: US-09-938-842A-1034
Effect score: 1071
Sequence: 1 atggcgacattcagaagct.....cagggtcatcgaaacctga 1071

Scoring table: OLIGO_NUC
Gapop 60.0, Gapext 60.0

Aligned: 22781392 seqs, 12152238056 residues

Word size: 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: EST*

- 1: em_estba.*
- 2: em_esthum.*
- 3: em_estin.*
- 4: em_estmu.*
- 5: em_estov.*
- 6: em_estpl.*
- 7: em_estro.*
- 8: em_htc.*
- 9: gb_est1.*
- 10: gb_est2.*
- 11: gb_htc.*
- 12: gb_est3.*
- 13: gb_est4.*
- 14: gb_est5.*
- 15: em_estfun.*
- 16: em_estom.*
- 17: em_gss_hum.*
- 18: em_gss_inv.*
- 19: em_gss_pln.*
- 20: em_gss_vrt.*
- 21: em_gss_fun.*
- 22: em_gss_mam.*
- 23: em_gss_mus.*
- 24: em_gss_pro.*
- 25: em_gss_rod.*
- 26: em_gss_phg.*
- 27: em_gss_vrl.*
- 28: gb_gss1.*
- 29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 334 | 31.2 | 373 | 10 | BE522768 |
| 2 | 320 | 29.9 | 486 | 9 | AV825375 |
| 3 | 244 | 22.8 | 378 | 10 | BE523004 |
| 4 | 224 | 20.9 | 388 | 14 | Z35068 |

| | | | | | |
|----|-----|------|-----|----|----------|
| 5 | 108 | 10.1 | 335 | 14 | T88453 |
| 6 | 38 | 3.5 | 586 | 28 | BH741132 |
| 7 | 38 | 3.5 | 798 | 28 | BH511920 |
| 8 | 35 | 3.3 | 656 | 28 | BH450567 |
| 9 | 32 | 3.0 | 600 | 28 | BH735681 |
| 10 | 24 | 2.2 | 318 | 9 | AW092131 |
| 11 | 24 | 2.2 | 481 | 9 | AI775374 |
| 12 | 24 | 2.2 | 610 | 9 | AI484214 |
| 13 | 24 | 2.2 | 642 | 9 | AW041016 |
| 14 | 24 | 2.2 | 670 | 10 | BG129822 |
| 15 | 23 | 2.1 | 222 | 10 | BB042901 |
| 16 | 23 | 2.1 | 222 | 10 | BB052744 |
| 17 | 23 | 2.1 | 222 | 10 | BB052810 |
| 18 | 23 | 2.1 | 224 | 10 | BB057092 |
| 19 | 23 | 2.1 | 226 | 10 | BB145694 |
| 20 | 23 | 2.1 | 232 | 10 | BB118640 |
| 21 | 23 | 2.1 | 233 | 10 | BB134002 |
| 22 | 23 | 2.1 | 307 | 10 | BB494924 |
| 23 | 23 | 2.1 | 381 | 13 | BY506397 |
| 24 | 23 | 2.1 | 390 | 13 | BY676112 |
| 25 | 23 | 2.1 | 395 | 13 | BY494054 |
| 26 | 23 | 2.1 | 415 | 14 | H36511 |
| 27 | 23 | 2.1 | 435 | 13 | BY583791 |
| 28 | 23 | 2.1 | 441 | 13 | BY436758 |
| 29 | 23 | 2.1 | 479 | 13 | BU635805 |
| 30 | 23 | 2.1 | 518 | 9 | AI484215 |
| 31 | 21 | 2.0 | 499 | 9 | AI795304 |
| 32 | 21 | 2.0 | 510 | 9 | AI746274 |
| 33 | 21 | 2.0 | 559 | 9 | AI668367 |
| 34 | 21 | 2.0 | 567 | 9 | AI665001 |
| 35 | 21 | 2.0 | 721 | 28 | BZ001843 |
| 36 | 21 | 2.0 | 747 | 13 | BQ865644 |
| 37 | 20 | 1.9 | 146 | 28 | BH617024 |
| 38 | 20 | 1.9 | 208 | 28 | AZ213090 |
| 39 | 20 | 1.9 | 278 | 28 | BZ377781 |
| 40 | 20 | 1.9 | 339 | 28 | BH657297 |
| 41 | 20 | 1.9 | 371 | 9 | AA645472 |
| 42 | 20 | 1.9 | 377 | 10 | BF649315 |
| 43 | 20 | 1.9 | 385 | 9 | AI506347 |
| 44 | 20 | 1.9 | 400 | 14 | CB699695 |
| 45 | 20 | 1.9 | 427 | 14 | CB795370 |

ALIGNMENTS

| | | | | | |
|------------|--|--------|------|--------|---------------------------|
| RESULT 1 | BE522768 | 373 bp | mrna | linear | EST 19-MAR-2001 |
| LOCUS | M28F8STM Arabidopsis developing seed | | | | Arabidopsis thaliana cDNA |
| DEFINITION | Clone M28F8 5', mRNA sequence. | | | | |
| ACCESSION | BE522768 | | | | |
| VERSION | BE522768.1 | | | | GI:9780746 |
| KEYWORDS | EST. | | | | |
| SOURCE | Arabidopsis thaliana (thale cress) | | | | |
| ORGANISM | Arabidopsis thaliana | | | | |
| REFERENCE | White, J.A., Todd, J.J., Newman, T., Focks, N., Girke, T., Martinez de Ibarra, O., Javorski, J.G., Ohlrogge, J., and Benning, C. 1999. A new set of Arabidopsis expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil Plant Physiol. 124 (4), 1582-1594 (2000) | | | | |
| AUTHORS | White, J.A., Todd, J.J., Newman, T., Focks, N., Girke, T., Martinez de Ibarra, O., Javorski, J.G., Ohlrogge, J., and Benning, C. | | | | |
| TITLE | A new set of Arabidopsis expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil | | | | |
| JOURNAL | Plant Physiol. 124 (4), 1582-1594 (2000) | | | | |
| MEDLINE | 20567808 | | | | |
| PUBMED | 11115876 | | | | |
| COMMENT | Contact: Benning, C Dept. of Biochemistry & Molecular Biology Michigan State University 224 Biochemistry, Michigan State University, East Lansing, MI 48824 , USA Tel: 517 355 1609 | | | | |

Fax: 517 353 9334

Email: benning@osu.edu

Michigan State University DNA Sequencing Facility Arabidopsis

Biological Resource Center, The Ohio State University, 309 Botany &

Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:

6142920603 TEL: 6142929371.

FEATURES

Location/Qualifiers

1. 373

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/strain="Columbia"

/db_xref="taxon:3702"

/clone="M28P8"

/tissue_type="seed"

/dev_stage="5-13 days after flowering"

/lab_host="E.coli"

/clone_lib="Arabidopsis developing seed"

/notes="Organ: Developing seed; Vector: pBluescript SK-"

Site1: EcoRI; Site2: XhoII"

73 a 118 c 82 g 100 t

BASE COUNT

ORIGIN

Query Match 31.2%; Score 334; DB 10; Length 373;

Best Local Similarity 100.0%; Pred. No. 6.9e-174;

Matches 334; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

536 GCGACGCGGTTTCAGCTTCCTCGGTTTAGCTCCAAATGCGCAGCAGCAAGATCCAC 595

Db

12 GCGACGCGGTTTCAGCTTCCTCGGTTTAGCTCCAAATGCGCAGCAGCAAGATCCAC 71

2Y

596 CTCGCAAGCTCTGGCATCTCACTGTGCTCAGCAACTCTGCGCAGCAAGATGTATC 655

Db

72 CTCGCAAGCTCTGGCATCTCACTGTGCTCAGCAACTCTGCGCAGCAAGATGTATC 131

2Y

656 CGATGTGGGCTATTCCATCAAGCAATGATTCGACGGTTCGAGCTTCTCTTGAATTC 715

Db

132 CGATGTGGGCTATTCCATCAAGCAATGATTCGACGGTTCGAGCTTCTCTTGAATTC 191

2Y

716 CACAATCGTGCTGCTCGTCAATCAGCCCTAGTATTAGCTTTCCGCGCGGCTGCTT 775

Db

192 CACAATCGTGCTGCTCGTCAATCAGCCCTAGTATTAGCTTTCCGCGCGGCTGCTT 251

2Y

776 CGCGCTGCTTACGTGCGGCTGTTCAACAGGCTTCCAGATGGCTAGACCACTCTT 835

Db

252 CGCGCTGCTTACGTGCGGCTGTTCAACAGGCTTCCAGATGGCTAGACCACTCTT 311

QY

836 TACAAGTTGTTCAGCAGCGGCTTTGTATCCGT 869

Db

312 TACAAGTTGTTCAGCAGCGGCTTTGTATCCGT 345

RESULT 2

AV825375

LOCUS AV825375 RAFL7 Arabidopsis thaliana cDNA clone RAFL07-08-P04 5',

mRNA sequence.

DEFINITION AV825375.1 GI:19867435

ACCESSION AV825375

VERSION AV825375.1

KEYWORDS Arabidopsis thaliana (thale cress)

SOURCE Arabidopsis thaliana

ORGANISM Arabidopsis thaliana

Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eutrosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 486)

REFERENCE Seki.M., Narusaka.M., Ishida.J., Kamiya.A., Satou.M., Nakajima.M.,

Ono.Y., Sakurai.T., Carninci.P., Kawai.J., Itoh.M., Ishii.Y.,

Arakawa.T., Shibata.K., Shinagawa.A., Muramatsu.M., Hayashizaki.Y.

and Shinozaki.K.

Large scale analysis of Arabidopsis full-length cDNA (2002b)

UNPUBLISHED

CONTACT: Morioaki Seki

Plant Functional Genomics Research Group

RIKEN Genomic Sciences Center

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: 81-298-36-4359

Fax: 81-298-36-9060

Email: msekig@rc.riken.go.jp

An Arabidopsis full-length cDNA library was constructed essentially

as reported previously (Seki et al., 1998). cDNA cleaved with BamHI

and XhoI was ligated to modified Lambda PUC-1 vector (Carninci et

al., submitted for publication) digested with BamHI and SalI. This

clone is in a modified pBluescript vector. Please visit our web

site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further

details.

Location/Qualifiers

1. 486

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/db_xref="taxon:3702"

/clone="RAFL07-08-P04"

/dev_stage="rosette plants"

/lab_host="DH10B"

/clone_lib="RAFL7"

/notes="Site 1: BamHI; Site 2: SalI; subjected to

cold-treated (1, 2, 5, 10, 24 hr)"

131 a 118 c 117 g 117 t 3 others

BASE COUNT

ORIGIN

Query Match 29.9%; Score 320; DB 9; Length 486;

Best Local Similarity 100.0%; Pred. No. 4.5e-166;

Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1 ATGCGCAATTCAGAGCTTGAGAGTTGCGGCAAGATCAAACTCTTAAGAGCGGTT 60

Db

102 ATGCGCAATTCAGAGCTTGAGAGTTGCGGCAAGATCAAACTCTTAAGAGCGGTT 161

QY

61 GATCTAACCATCATCAACGGCTCAGAAAGCTCGAAACTTCAAGACCTTCCAAAGTAAT 120

Db

162 GATCTAACCATCATCAACGGCTCAGAAAGCTCGAAACTTCCAAAGTAAT 221

QY

121 CCCACAGTGTCTCGAGCCCAAGGGGAGCGGTGATCGCGTTCGATGCTCTTTA 180

Db

222 CCCACAGTGTCTCGAGCCCAAGGGGAGCGGTGATCGCGTTCGATGCTCTTTA 281

QY

181 GCTCCACCGCTTCGACAGGACCACTTGAAGAGAGCTTCGACTTAAGACCGTCAACG 240

Db

282 GCTCCACCGCTTCGACAGGACCACTTGAAGAGAGCTTCGACTTAAGACCGTCAACG 341

QY

241 AAGTTGAAGAGAGGAGAGATACCGATCCCTGCCACGTGTCGGCTAGGATTTT 300

Db

342 AAGTTGAAGAGAGGAGAGATACCGATCCCTGCCACGTGTCGGCTAGGATTTT 401

QY

301 CAATTAACCTGAGAGTTAGG 320

Db

402 CAATTAACCTGAGAGTTAGG 421

RESULT 3

BE523004

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana

Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eutrosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 378)

REFERENCE

AUTHORS

TITLE

COMMENT

BE523004 M1B6STM Arabidopsis developing seed Arabidopsis thaliana cDNA

clone M31B6 5', mRNA sequence.

378 bp mRNA linear EST 19-MAR-2001

clone M31B6 5', mRNA sequence.

BE523004.1 GI:9780982

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eutrosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 378)

REFERENCE

AUTHORS

TITLE

COMMENT

A new set of Arabidopsis expressed sequence tags from developing

seeds. The metabolic pathway from carbohydrates to seed oil

Email: 22313tcn@bm.cl.msu.edu
 Seq primer: T7 dye primer.
 Location/Qualifiers
 1..335
 source

/organism="Arabidopsis thaliana"

/mol_type="rRNA"

/strain="var columbia"

/db_xref="taxon:3702"

/clone="156J17"

/clone_lib="Lambda-PRL2"

/note="Vector: lambda Zip-Lox; Site 1: Sal; Site 2: Not; Lambda PRL2 is a cDNA library derived from equal quantities of 4 pools of mRNA. The mRNA sources were 1) 7 day germinated etiolated seedlings; 2) tissue culture grown roots; 3) staged plants half with 24 hour light cycle, half on 16 hr light, 8 hour dark- rosettes; 4) same plants as 3 but aerial tissue (stems, flowers and siliques). The vector is BRU's lambda Zip-Lox. The cDNA inserts were directionally cloned with Sal-Not arms using oligo dt primed cDNA."

BASE COUNT 96 a 75 c 82 g 69 t 13 others

ORIGIN

Query Match 10.1%; Score 108; DB 14; Length 335;

Best Local Similarity 100.0%; Pred. No. 2.7e-48; Length 335;

Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 42 TCAACTCTAAGAGCGGTTGATCTAACCATCATCAACGGCGTCAGAAAGTCGAAACTTC 101

Db 40 TCAACTCTAAGAGCGGTTGATCTAACCATCATCAACGGCGTCAGAAAGTCGAAACTTC 99

2Y 102 AAGACCTTCCAAAGTAATCCACAGTGTCTGAGCCCAAGGCGGA 149

Db 100 AAGACCTTCCAAAGTAATCCACAGTGTCTGAGCCCAAGGCGGA 147

RESULT 6

3B741132/c

LOCUS BH741132 586 bp DNA linear GSS 25-FEB-2002
DEFINITION gt27g10.g1 BoBuds01 Brassica oleracea genomic clone gt27g10 5',
genomic survey sequence.

ACCESSION BH741132

VERSION BH741132.1 GI:18875745

KEYWORDS GSS.

SOURCE Brassica oleracea

ORGANISM Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 586)

REFERENCE Katari, M., O'Shaughnessy, A., Palmer, L., Bahret, A., Baker, J., Balijsa

V., Cunniss, D.M., Katzenberger, F., King, L., Kirchoff, K., Kuit, K.,

Miller, B., Muller, S., Nascimento, L., Preston, R., Santos, L., Shah, R.,

Zutavern, T., Dedhia, N., Rabinowicz, P.D. and McCombie, W.R.

Whole Genome Shotgun Reads from Brassica oleracea (2002b)

Unpublished

TITLE

JOURNAL

COMMENT

Contact: W. Richard McCombie

Lita Annenberg Hazen Genome Sequencing Center

Cold Spring Harbor Laboratory

PO Box 100, Cold Spring Harbor, NY 11724, USA

Tel: 516 367 8884

Fax: 516 367 8874

Email: mcombie@cshl.org

Plate: gt27 row: g column: 10

Seq primer: -21UnivRev

Class: shotgun

High quality sequence stop: 586.

Location/Qualifiers

1. 586

/organism="Brassica oleracea"

/mol_type="genomic DNA"

/db_xref="taxon:3712"

/clone="gt27g10"

/clone_lib="BoBuds01"

/note="Vector: M13 for .x reads, pBluescript for .b and .g

reads; Site 1: EcoRV; Whole genome shotgun library from

flowering buds. DNA was purified from a crude nuclear prep

FEATURES

source

RESULT 8

BH450567/c

LOCUS BH450567 656 bp DNA linear GSS 12-DEC-2001

DEFINITION BOGDQ41TR BOGD Brassica oleracea genomic clone BOGDQ41, genomic

survey sequence.

ACCESSION BH450567

VERSION BH450567.1 GI:17636278

KEYWORDS GSS.

using Brassica oleracea TC1000DH3 buds provided by Thomas Osborn at the University of Wisconsin. Genomic DNA provided by Pablo Rabinowicz (CSHL) and shotgun library prepared in McCombie Lab."

BASE COUNT 162 a 120 c 131 g 173 t

ORIGIN

Query Match 3.5%; Score 38; DB 28; Length 586;

Best Local Similarity 100.0%; Pred. No. 2.6e-09;

Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 262 AGGATACGGATGCTGCCACGTGTGGCGCTAGGATTTT 299

Db 41 AGGATACGGATGCTGCCACGTGTGGCGCTAGGATTTT 4

RESULT 7

BH511920/c

LOCUS BH511920 798 bp DNA linear GSS 13-DEC-2001

DEFINITION BOHGA14TR BOHG Brassica oleracea genomic clone BOHGA14, genomic

survey sequence.

ACCESSION BH511920

VERSION BH511920.1 GI:17720010

KEYWORDS GSS.

SOURCE Brassica oleracea

ORGANISM Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 798)

REFERENCE Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.

Whole genome shotgun sequencing of Brassica oleracea

Unpublished

Other GSSs: BOHGA14TR

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TF

Class: sheared ends.

Location/Qualifiers

1. 798

/organism="Brassica oleracea"

/mol_type="genomic DNA"

/strain="TC1000DH3"

/db_xref="taxon:3712"

/clone="BOHGA14"

/clone_lib="BOHG"

/note="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared

genomic DNA inserted into pHOS1 using BstXI linkers"

161 a 210 c 225 g 202 t

BASE COUNT

ORIGIN

Query Match 3.5%; Score 38; DB 28; Length 798;

Best Local Similarity 100.0%; Pred. No. 2.8e-09;

Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 262 AGGATACGGATGCTGCCACGTGTGGCGCTAGGATTTT 299

Db 720 AGGATACGGATGCTGCCACGTGTGGCGCTAGGATTTT 683

SOURCE
ORGANISM
Brassica oleracea
Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 656)

REFERENCE
TOWN, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished
JOURNAL
COMMENT
Other GSSs: BOGDQ41TF
Contact: Chris Town

TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.

FEATURES
source
Location/Qualifiers

1..656
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TOL000DH3"
/db_xref="taxon:3712"
/clone="BOGDQ41"
/clone_lib="BOGD"
/note="Vector: pBstXI; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into pBstXI using BstXI linkers"

142 a 167 c 190 g 157 t

BASE COUNT

Query Match 3.3%; Score 35; DB 28; Length 656;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 382 CCGCCACGGTAGCGGAACGGTTCGCCCATCGC 416
D 563 CCGCCACGGTAGCGGAACGGTTCGCCCATCGC 529

RESULT 9
H735681
OCUS
EFINITION
BOHTG58TR BO 2.3 KB Brassica oleracea genomic clone BOHTG58,
genomic survey sequence.

CESSION
H735681
ERSON
H735681.1 GI:18841076
EYWORDS
GSS.

SOURCE
ORGANISM
Brassica oleracea
Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 600)

REFERENCE
TOWN, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished
JOURNAL
COMMENT
Other GSSs: BOHTG58TF
Contact: Chris Town

TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.

FEATURES
source
Location/Qualifiers
1..600
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TOL000DH3"

/db_xref="taxon:3712"

/clone="BOHTG58"

/clone_lib="BO 2.3 KB"

/note="Vector: pBstXI; Site 1: BstXI; 2-3 kb sheared

genomic DNA inserted into pBstXI using BstXI linkers"

159 a 126 c 160 g 155 t

BASE COUNT
ORIGIN

Query Match 3.0%; Score 32; DB 28; Length 600;

Best Local Similarity 100.0%; Pred. No. 5.6e-06;

Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 ACTCGAGAGTTAGGTCAACAATCCGACGGCGA 338

DB 119 ACTCGAGAGTTAGGTCAACAATCCGACGGCGA 150

RESULT 10

AW092131

LOCUS

DEFINITION

AW092131

VERSION

AW092131.1

KEYWORDS

SOURCE

ORGANISM

Lycopersicon esculentum (tomato)

Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

asterids; lamiales; Solanales; Solanaceae; Solanum; Lycopersicon.

1 (bases 1 to 318)

REFERENCE

AUTHORS

D'Ascenzo, M., He, X., Lyman, J., Holt, I.E., Liang, F., Upton, J.,

Ronning, C.M., Craven, M.B., Fujii, C.Y., Bowman, C.L., Nierman, W.,

Fraser, C.M., Venter, J.C., Martin, G.B., Tanksley, S.D. and Giovannoni

J.

Generation of ESTs from tomato leaf tissue

Unpublished

Contact: CUGI

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

5 prime sequence.

FEATURES

source

1..318

/organism="Lycopersicon esculentum"

/mol_type="mRNA"

/cultivar="Rio Grande PtoR"

/db_xref="taxon:4081"

/clone="cLET17P7"

/tissue_type="leaf"

/dev_stage="4-6 week old plants"

/lab_host="XLA-Blue MRF"

/clone_lib="tomato mixed elicitor, BTL"

/note="Vector: pBlueScript SK(-); Site 1: EcoRI; Site 2:

XhoI; cLET - inoculated with a variety of disease response

elicitors. Plants exposed to 2,6 dichloroisonicotinic

acid, BTH, jasmonic acid, ethylene, fenthion, EIX,

okadaic acid, or systemin prior to tissue harvest. EcoRI

site was destroyed during cloning."

58 a 78 c 82 g 100 t

BASE COUNT

ORIGIN

Query Match 2.2%; Score 24; DB 9; Length 318;

Best Local Similarity 100.0%; Pred. No. 0.13;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 761 CCGCCGCCCTCTTCGCGCGTGT 784

DB 36 CCGCCGCCCTCTTCGCGCGTGT 59

RESULT 11

AI775374
LOCUS EST256474 tomato resistant, Cornell Lycopersicon esculentum cDNA
DEFINITION
ACCESSION AI775374
VERSION AI775374.1 GI:5273415
KEYWORDS
SOURCE
ORGANISM Lycopersicon esculentum (tomato)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
1 (bases 1 to 481)
D'Ascenzo, M., He, X., Lyman, J., Matern, A.L., Vision, T., Holt, I.E.,
Liang, F., Upton, J., Ronning, C.M., Craven, M.B., Fujii, C.Y., Bowman
C.L., Nierman, W., Fraser, C.M., Venter, J.C., Tanksley, S.D.,
Giovannoni, J.J., and Martin, G.B.
Generation of ESTs from Pseudomonas resistant tomato
Unpublished
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.
Location/Qualifiers
1..481
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="R11-12 (35S::Pto in Rio Grande x Money Maker)"
/db_xref="taxon:4081"
/clone="CLER15G18"
/tissue_type="leaf"
/dev_stage="4-week old"
/lab_host="SOLR"
/clone_lib="tomato resistant, Cornell"
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; cLER - Tomato Pseudomonas Resistant EST Library.
Directionally cloned cDNAs inserted into pBluescript SK(-
) at 5' end with EcoRI and 3' end with XhoI site."
BASE COUNT 109 a 114 c 110 g 148 t
ORIGIN
Query Match 2.2%; Score 24; DB 9; Length 481;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
2y 761 CCGCCCGCTGCTTCGCGTCTGT 784
80 CCGCCCGCTGCTTCGCGTCTGT 103
EST.
AI484214 610 bp mRNA linear EST 18-MAY-2001
LOCUS
DEFINITION EST249484 tomato resistant, Cornell Lycopersicon esculentum cDNA
ACCESSION AI484214
VERSION AI484214.1 GI:4379585
KEYWORDS
SOURCE
ORGANISM Lycopersicon esculentum (tomato)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
1 (bases 1 to 610)
D'Ascenzo, M., He, X., Lyman, J., Matern, A.L., Vision, T., Holt, I.E.,
Liang, F., Upton, J., Ronning, C.M., Craven, M.B., Fujii, C.Y., Bowman
C.L., Nierman, W., Fraser, C.M., Venter, J.C., Tanksley, S.D.,
Giovannoni, J.J., and Martin, G.B.
Generation of ESTs from Pseudomonas resistant tomato
Unpublished
Other_ESTs: EST249485

Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>.
Location/Qualifiers
1..610
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="R11-12 (35S::Pto in Rio Grande x Money Maker)"
/db_xref="taxon:4081"
/clone="CLER1A9"
/tissue_type="leaf"
/dev_stage="4-week old"
/lab_host="SOLR"
/clone_lib="tomato resistant, Cornell"
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; cLER - Tomato Pseudomonas Resistant EST Library.
Directionally cloned cDNAs inserted into pBluescript SK(-
) at 5' end with EcoRI and 3' end with XhoI site."
BASE COUNT 194 a 138 c 128 g 150 t
ORIGIN
Query Match 2.2%; Score 24; DB 9; Length 610;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 761 CCGCCCGCTGCTTCGCGTCTGT 784
Db 602 CCGCCCGCTGCTTCGCGTCTGT 579
EST.
AW041016 642 bp mRNA linear EST 18-MAY-2001
LOCUS
DEFINITION EST23880 tomato mixed elicitor, BTI Lycopersicon esculentum cDNA
ACCESSION AW041016
VERSION AW041016.1 GI:5899770
KEYWORDS
SOURCE
ORGANISM Lycopersicon esculentum (tomato)
Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
1 (bases 1 to 642)
D'Ascenzo, M., He, X., Lyman, J., Holt, I.E., Liang, F., Upton, J.,
Ronning, C.M., Craven, M.B., Fujii, C.Y., Bowman, C.L., Nierman, W.,
Fraser, C.M., Venter, J.C., Martin, G.B., Tanksley, S.D. and Giovannoni
J.J.
Generation of ESTs from tomato leaf tissue
Unpublished
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.
Location/Qualifiers
1..642
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="Rio Grande PtoR"
/db_xref="taxon:4081"
/clone="cLER1G21"
/tissue_type="leaf"
/dev_stage="4-6 week old plants"
/lab_host="X11-Blue MRF"
/clone_lib="tomato mixed elicitor, BTI"
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; cLER - inoculated with a variety of disease response
elicitors. Plants exposed to 2,6 dichloroisonicotinic
acid, BTH, Jasmonic acid, ethylene, fenthion, EIX,

ORIGIN

Query Match 2.1%; Score 23; DB 10; Length 222;
 Best Local Similarity 100.0%; Pred. No. 0.41; 0; Indels 0; Gaps 0;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 112 CAAGTAAATCCACAGTGAGTCT 134
 |||||
 Db 149 CAAGTAAATCCACAGTGAGTCT 171

Search completed: February 2, 2004, 05:01:17
 Job time : 2490.95 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

M nucleic - nucleic search, using sw model,
un on: February 2, 2004, 02:56:51 ; Search time 71.3221 Seconds
(without alignments)
6627.983 Million cell updates/sec

title: US-09-938-842A-1034
effect score: 1071
equence: 1 atggcgacaattcagaagct.....cacggtcatgaaccactga 1071

coring table: OLIGO NUC
Gapop 60.0, Gapext 60.0

searched: 569978 seqs, 220691566 residues

ord size : 0
otal number of hits satisfying chosen parameters: 1139956

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Listing first 45 summaries

atabase : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|---------|---------------------|--------------------|
| 1 | 19 | 1.8 | 34094 | US-09-292-034-1 | Sequence 1, Appli |
| 2 | 18 | 1.7 | 162450 | US-09-345-882-1 | Sequence 1, Appli |
| 3 | 18 | 1.7 | 1664976 | US-08-916-421B-1 | Sequence 1, Appli |
| 4 | 17 | 1.6 | 447 | US-09-328-352-186 | Sequence 186, App |
| 5 | 17 | 1.6 | 516 | US-08-510-878-2 | Sequence 2, Appli |
| 6 | 17 | 1.6 | 748 | US-08-510-878-3 | Sequence 3, Appli |
| 7 | 17 | 1.6 | 867 | US-09-252-991A-6878 | Sequence 6878, Ap |
| 8 | 17 | 1.6 | 882 | US-09-252-991A-6792 | Sequence 6792, Ap |
| 9 | 17 | 1.6 | 1353 | US-08-611-280-1 | Sequence 1, Appli |
| 10 | 17 | 1.6 | 1353 | US-09-195-940-1 | Sequence 1, Appli |
| 11 | 17 | 1.6 | 1353 | US-09-562-466-1 | Sequence 1, Appli |
| 12 | 17 | 1.6 | 1960 | US-09-553-867A-42 | Sequence 42, Appli |
| 13 | 17 | 1.6 | 3758 | US-08-323-477-1 | Sequence 4, Appli |
| 14 | 17 | 1.6 | 12537 | US-08-611-280-4 | Sequence 4, Appli |
| 15 | 17 | 1.6 | 12537 | US-09-195-940-4 | Sequence 4, Appli |
| 16 | 17 | 1.6 | 12537 | US-09-562-466-4 | Sequence 4, Appli |
| 17 | 17 | 1.6 | 1664976 | US-08-916-421B-1 | Sequence 1, Appli |
| 18 | 17 | 1.6 | 4403765 | US-09-103-840A-2 | Sequence 2, Appli |
| 19 | 17 | 1.6 | 4411529 | US-09-103-840A-1 | Sequence 1, Appli |
| 20 | 16 | 1.5 | 268 | US-08-444-818-21 | Sequence 21, Appli |
| 21 | 16 | 1.5 | 306 | US-08-411-913-8 | Sequence 8, Appli |
| 22 | 16 | 1.5 | 307 | US-08-444-818-143 | Sequence 143, App |
| 23 | 16 | 1.5 | 438 | US-09-255-000-8 | Sequence 8, Appli |
| 24 | 16 | 1.5 | 477 | US-07-853-985A-7 | Sequence 7, Appli |
| 25 | 16 | 1.5 | 477 | US-07-681-703B-7 | Sequence 7, Appli |
| 26 | 16 | 1.5 | 477 | US-08-184-236-7 | Sequence 7, Appli |
| 27 | 16 | 1.5 | 477 | US-08-407-410B-7 | Sequence 7, Appli |

C 28 16 1.5 477 2 US-08-485-500-7 Sequence 7, Appli
C 29 16 1.5 477 5 PCT-US91-02370-7 Sequence 7, Appli
C 30 16 1.5 477 5 PCT-US94-04174-7 Sequence 7, Appli
C 31 16 1.5 489 4 US-09-328-352-1768 Sequence 1768, Ap
C 32 16 1.5 493 4 US-09-364-206-5 Sequence 5, Appli
C 33 16 1.5 557 4 US-09-364-206-6 Sequence 6, Appli
C 34 16 1.5 558 1 US-07-853-985A-9 Sequence 9, Appli
C 35 16 1.5 558 1 US-07-681-703B-9 Sequence 9, Appli
C 36 16 1.5 558 1 US-08-184-236-9 Sequence 9, Appli
C 37 16 1.5 558 2 US-08-407-410B-9 Sequence 9, Appli
C 38 16 1.5 558 2 US-08-485-500-9 Sequence 9, Appli
C 39 16 1.5 558 5 PCT-US91-02370-9 Sequence 9, Appli
C 40 16 1.5 558 5 PCT-US94-04174-9 Sequence 9, Appli
C 41 16 1.5 701 3 US-08-998-416-908 Sequence 908, App
C 42 16 1.5 721 4 US-09-205-258-141 Sequence 141, App
C 43 16 1.5 800 3 US-08-444-818-31 Sequence 31, Appli
C 44 16 1.5 816 1 US-08-350-884-83 Sequence 83, Appli
C 45 16 1.5 816 1 US-08-440-548-83 Sequence 83, Appli

ALIGNMENTS

RESULT 1
US-09-292-034-1
; Sequence 1, Application US/09292034
; Patent No. 6492343
; GENERAL INFORMATION:
; APPLICANT: Reddy, P. Seshidhar
; APPLICANT: Bablu, Suresh
; APPLICANT: Tikoo, Lorne
; TITLE OF INVENTION: PORCINE ADENOVIRUS TYPE 3 GENOME
; FILE REFERENCE: 293102002400
; CURRENT APPLICATION NUMBER: US/09/292,034
; CURRENT FILING DATE: 1999-04-14
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: fastseq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 34094
; TYPE: DNA
; ORGANISM: Porcine Adenovirus Type 3
; FEATURE:
US-09-292-034-1

Query Match 1.8%; Score 19; DB 4; Length 34094;
Best Local Similarity 100.0%; Pred.No.4;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 850 AGCAGCGGCTTTGTATCCG 868
Db 21527 AGCAGCGGCTTTGTATCCG 21545

RESULT 2
US-09-345-882-1
; Sequence 1, Application US/09345882
; Patent No. 6399373
; GENERAL INFORMATION:
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP-7)
; TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP-7)
; FILE REFERENCE: GENSET.031A
; CURRENT APPLICATION NUMBER: US/09/345,882
; CURRENT FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: US 60/091,315
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/111,909
; PRIOR FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 140
; SOFTWARE: Patent.pm
; SEQ ID NO 1
; LENGTH: 162450
; TYPE: DNA

```
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: allele
LOCATION: 146345
OTHER INFORMATION: 5-124-273 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 88073
OTHER INFORMATION: 5-127-261 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 90842
OTHER INFORMATION: 99-1437-325 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 93714
OTHER INFORMATION: 5-128-60 : polymorphic base deletion of GT
FEATURE:
NAME/KEY: allele
LOCATION: 97122
OTHER INFORMATION: 99-1442-224 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 97152
OTHER INFORMATION: 5-129-144 : polymorphic base deletion of T
FEATURE:
NAME/KEY: allele
LOCATION: 99098
OTHER INFORMATION: 5-130-257 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 99117
OTHER INFORMATION: 5-130-276 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 103806
OTHER INFORMATION: 5-131-395 : polymorphic base A or T
FEATURE:
NAME/KEY: allele
LOCATION: 106940
OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 108106
OTHER INFORMATION: 5-135-155 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 108149
OTHER INFORMATION: 5-135-198 : polymorphic base insertion of GTT
FEATURE:
NAME/KEY: allele
LOCATION: 108308
OTHER INFORMATION: 5-135-357 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 108471
OTHER INFORMATION: 5-136-174 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 134134
OTHER INFORMATION: 5-140-120 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 134362
OTHER INFORMATION: 5-140-348 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 134374
OTHER INFORMATION: 5-140-361 : polymorphic base insertion of CA
FEATURE:
NAME/KEY: allele
LOCATION: 146328
OTHER INFORMATION: 5-143-84 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 146345
OTHER INFORMATION: 5-143-101 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 150329
OTHER INFORMATION: 5-145-24 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 160031
OTHER INFORMATION: 5-148-352 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID30
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID51
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID31
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID52
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID49
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID70
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID53
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID50
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID71
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID33
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID34
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID55
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID35
FEATURE:
```

NAME/KEY: allele
LOCATION: 98094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID56
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID36
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID57
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID37
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID58
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID38
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID59
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID39
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID60
FEATURE:

Query Match 1.7%; Score 18; DB 4; Length 162450;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 695 TCGAGCTTCTCTCTGA 712

b 35413 TCGAGCTTCTCTCTGA 35430

RESULT 3

US-08-916-421B-1/c
Sequence 1, Application US/08916421B
Patent No. 6503729
GENERAL INFORMATION:
APPLICANT: Bult et al.
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
Patent No. 6503729
TITLE OF INVENTION: jannaschii
FILE REFERENCE: PB275
CURRENT APPLICATION NUMBER: US/08/916,421B
CURRENT FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/024,428
PRIOR FILING DATE: 1996-08-22
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1664976
TYPE: DNA
ORGANISM: Methanococcus jannaschii
FEATURE:
NAME/KEY: misc_feature
LOCATION: (28222)..(28222)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (28257)..(28258)
OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc_feature
LOCATION: (84773)..(84773)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (84808)..(84808)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (84812)..(84812)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (103998)..(103998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234814)..(234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (319226)..(319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559241)..(559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature

LOCATION: (600992)..(600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349491)..(1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1470091)..(1470091)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1569020)..(1569020)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1602912)..(1602912)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1603734)..(1603734)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1637998)..(1637998)

OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1664854)..(1664855)
OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

Query Match 1.7%; Score 18; DB 4; Length 1664976;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 CAATGCTTTTACCTCCAC 187
DB 1522815 CAATGCTTTTACCTCCAC 1522798

RESULT 4

US-09-328-352-186
Sequence 186, Application US/09328352
Patent No. 6562958

GENERAL INFORMATION:

APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 186
LENGTH: 447
TYPE: DNA
ORGANISM: Acinetobacter baumannii
US-09-328-352-186

Query Match 1.6%; Score 17; DB 4; Length 447;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 951 TTCATCGTCATTCGCAA 967
DB 326 TTCATCGTCATTCGCAA 342

RESULT 5

US-08-510-878-2
Sequence 2, Application US/08510878
Patent No. 5776771

GENERAL INFORMATION:

APPLICANT: Yu, Fujio
APPLICANT: Kato, Mami
TITLE OF INVENTION: A KANAMYCIN RESISTANCE GENE DERIVED FROM
TITLE OF INVENTION: MICROORGANISMS OF THE GENUS RHODOCOCUS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch and Birch
STREET: P.O. Box 747
CITY: Falls Church
STATE: VA
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/510,878
FILING DATE: 03-AUG-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Weiner, Marc S
REGISTRATION NUMBER: 32,181
REFERENCE/DOCKET NUMBER: 1254-121
TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 516 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
S-08-510-878-2

Query Match 1.6%; Score 17; DB 1; Length 516;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 439 AAAATCCCGACGACGAC 455
b 120 AAAATCCCGACGACGAC 136

RESULT 6

S-08-510-878-3
Sequence 3, Application US/08510878
Patent No. 5776771

GENERAL INFORMATION:

APPLICANT: Yu, Fujio
APPLICANT: Kato, Mami
TITLE OF INVENTION: A KANAMYCIN RESISTANCE GENE DERIVED FROM
TITLE OF INVENTION: MICROORGANISMS OF THE GENUS RHODOCYCUS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch and Birch
STREET: P.O. Box 747
CITY: Falls Church
STATE: VA
COUNTRY: USA
ZIP: 22040-0747

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/510,878
FILING DATE: 03-AUG-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Weiner, Marc S
REGISTRATION NUMBER: 32,181
REFERENCE/DOCKET NUMBER: 1254-121
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
LENGTH: 748 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
S-08-510-878-3

Query Match 1.6%; Score 17; DB 1; Length 748;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 439 AAAATCCCGACGACGAC 455
b 284 AAAATCCCGACGACGAC 300

RESULT 7

US-09-252-991A-6878
Sequence 6878, Application US/09252991A
Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 6878
LENGTH: 867
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6878

Query Match 1.6%; Score 17; DB 4; Length 867;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 CGCGGATTATAGCGGCC 387
DB 50 CGCGGATTATAGCGGCC 66

RESULT 8

US-09-252-991A-6792/c
Sequence 6792, Application US/09252991A
Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 6792
LENGTH: 882
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6792

Query Match 1.6%; Score 17; DB 4; Length 882;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 CGCGGATTATAGCGGCC 387
DB 82 CGCGGATTATAGCGGCC 66

RESULT 9

US-08-611-280-1/c
Sequence 1, Application US/08611280
Patent No. 5891666

GENERAL INFORMATION:

APPLICANT: Matsuyama, Toshifumi
APPLICANT: Grossman, Alex
APPLICANT: Richardson, Christopher D.
TITLE OF INVENTION: NOVEL GENES ENCODING LSIRF POLYPEPTIDES

NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Canada Inc.
STREET: 6733 Mississauga Road, Suite 303
CITY: Mississauga
STATE: Ontario
COUNTRY: Canada
ZIP: L5N 6J8
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/611,280
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Oleski, Nancy A.
REGISTRATION NUMBER: 34,688
REFERENCE/DOCKET NUMBER: A-338A
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1353 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-611-280-1
Query Match 1.6%; Score 17; DB 2; Length 1353;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 622 GTGGCTCAGCACTTCT 638
DB 930 GTGGCTCAGCACTTCT 914
RESULT 10
US-09-195-940-1/c
; Sequence 1, Application US/09195940
; Patent No. 6258935
; GENERAL INFORMATION:
; APPLICANT: Matsuyama, Toshifumi
; APPLICANT: Grossman, Alex
; TITLE OF INVENTION: NOVEL GENES ENCODING LSIRF POLYPEPTIDES
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Canada Inc.
; STREET: 6733 Mississauga Road, Suite 303
; CITY: Mississauga
; STATE: Ontario
; COUNTRY: Canada
; ZIP: L5N 6J8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/195,940
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/611,280
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Oleski, Nancy A.
; REGISTRATION NUMBER: 34,688
; REFERENCE/DOCKET NUMBER: A-338A
; INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1353 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-611-280-1

SEQUENCE CHARACTERISTICS:
LENGTH: 1353 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-195-940-1
Query Match 1.6%; Score 17; DB 3; Length 1353;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 622 GTGGCTCAGCACTTCT 638
DB 930 GTGGCTCAGCACTTCT 914
RESULT 11
US-09-562-466-1/c
; Sequence 1, Application US/09562466
; Patent No. 6369202
; GENERAL INFORMATION:
; APPLICANT: Matsuyama, Toshifumi
; APPLICANT: Grossman, Alex
; TITLE OF INVENTION: NOVEL GENES ENCODING LSIRF POLYPEPTIDES
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Canada Inc.
; STREET: 6733 Mississauga Road, Suite 303
; CITY: Mississauga
; STATE: Ontario
; COUNTRY: Canada
; ZIP: L5N 6J8
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/562,466
; FILING DATE: 01-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/195,940
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Oleski, Nancy A.
; REGISTRATION NUMBER: 34,688
; REFERENCE/DOCKET NUMBER: A-338A
; INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1353 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-562-466-1
Query Match 1.6%; Score 17; DB 4; Length 1353;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 622 GTGGCTCAGCACTTCT 638
DB 930 GTGGCTCAGCACTTCT 914
RESULT 12
US-09-553-867A-42/c
; Sequence 42, Application US/09553867A
; Patent No. 6476188

GENERAL INFORMATION:

APPLICANT: Young, Michael W
APPLICANT: Kloss, Brian
APPLICANT: Blau, Justin
APPLICANT: Price, Jeffrey
APPLICANT: Takahashi, Joseph S.
APPLICANT: Philip, Lowrey L.
TITLE OF INVENTION: A NOVEL CLOCK GENE AND METHODS OF USE THEREOF
FILE REFERENCE: 600-1-221NCP
CURRENT APPLICATION NUMBER: US/09/553.867A
PRIORITY FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: 60/090,068
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 09/335,983
PRIOR FILING DATE: 1999-06-18
NUMBER OF SEQ ID NOS: 42
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 42
LENGTH: 1960
TYPE: DNA
ORGANISM: Rattus norvegicus
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Primer.
S-09-553-867A-42

Query Match 1.6%; Score 17; DB 4; Length 1960;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 761 CCGCGCGCGTGTCTCG 777
D 101 CCGCGCGCGTGTCTCG 85

RESULT 13

S-08-323-477-1
Sequence 1, Application US/08323477
Patent No. 6086896
GENERAL INFORMATION:
APPLICANT: Sparling, P. Frederick
APPLICANT: Thompson, Stuart
TITLE OF INVENTION: ANTIGENIC IRON REPRESSIBLE PROTEINS FROM
TITLE OF INVENTION: N. MENINGITIDIS RELATED TO THE HEMOLYSIN FAMILY OF TOXINS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: ImClone Systems Incorporated
STREET: 180 Varick Street
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10014
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/323.477
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/920.963
FILING DATE: 28-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Feit, Irving N.
REGISTRATION NUMBER: 28,601
REFERENCE/DOCKET NUMBER: SPA-2-2P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-645-1405
TELEFAX: 212-645-2054
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3758 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Neisseria meningitidis
FEATURE:
NAME/KEY: CDS
LOCATION: 298..3645
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 298..3642
US-08-323-477-1

Query Match 1.6%; Score 17; DB 3; Length 3758;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1003 GAGATATACGAGAAACA 1019
DB 590 GAGATATACGAGAAACA 606

RESULT 14

US-08-611-280-4/c
Sequence 4, Application US/08611280
Patent No. 5891666
GENERAL INFORMATION:
APPLICANT: Matsuyama, Toshifumi
APPLICANT: Grossman, Alex
APPLICANT: Richardson, Christopher D.
TITLE OF INVENTION: NOVEL GENES ENCODING LSIRP POLYPEPTIDES
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Canada Inc.
STREET: 6733 Mississauga Road, Suite 303
CITY: Mississauga
STATE: Ontario
COUNTRY: Canada
ZIP: L5N 6J8
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/611.280
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Oleski, Nancy A.
REGISTRATION NUMBER: 34,688
REFERENCE/DOCKET NUMBER: A-338A
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 12537 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-611-280-4

Query Match 1.6%; Score 17; DB 2; Length 12537;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 622 GTGGCTCAGCACTTCT 638
DB 8349 GTGGCTCAGCACTTCT 8333

RESULT 15
 US-09-195-940-4/c
 ; Sequence 4, Application US/09195940
 ; Patent No. 6258935
 ; GENERAL INFORMATION:
 ; APPLICANT: Matsuyama, Toshifumi
 ; APPLICANT: Grossman, Alex
 ; APPLICANT: Richardson, Christopher D.
 ; TITLE OF INVENTION: NOVEL GENES ENCODING LSIRF POLYPEPTIDES
 ; NUMBER OF SEQUENCES: 25
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Amgen Canada Inc.
 ; STREET: 6733 Mississauga Road, Suite 303
 ; CITY: Mississauga
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: L5N 6J8
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/195,940
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/611,280
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Olecki, Nancy A.
 ; REGISTRATION NUMBER: 34,688
 ; REFERENCE/DOCKET NUMBER: A-338A
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 12537 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-09-195-940-4

Query Match 1.6%; Score 17; DB 3; Length 12537;
 Best Local Similarity 100.0%; Pred No. 44;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 622 GTGGCTCAGCACTTCT 638
 |||||
 Db 8349 GTGGCTCAGCACTTCT 8333

Search completed: February 2, 2004, 07:58:50
 Job time : 89.3221 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

M nucleic - nucleic search, using sw model

un on: February 2, 2004, 03:00:32 ; Search time 801.301 Seconds
(without alignments)
4871.226 Million cell updates/sec

itle: US-09-938-842A-1034
erfect score: 1071
equence: 1 atggcgacaattcagagct.....cagggctcagaaccactga 1071

coring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

earched: 2434939 seqs, 1822278265 residues

ord size : 0

otal number of hits satisfying chosen parameters: 4869878

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Listing first 45 summaries

atabase : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubnpa/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubnpa/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubnpa/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubnpa/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubnpa/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubnpa/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubnpa/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubnpa/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubnpa/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubnpa/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubnpa/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubnpa/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubnpa/US09_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubnpa/US10A_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubnpa/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubnpa/US10_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/1/pubnpa/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/1/pubnpa/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---------------------|
| 1 | 1071 | 100.0 | 1071 | 10 | US-09-938-842A-1034 |
| 2 | 1071 | 100.0 | 1071 | 12 | US-09-938-842A-1034 |
| 3 | 289 | 27.0 | 460 | 10 | US-09-938-842A-1034 |
| 4 | 286 | 26.7 | 453 | 9 | US-09-938-842A-1034 |
| 5 | 185 | 17.3 | 185 | 9 | US-09-938-842A-1034 |
| 6 | 23 | 2.1 | 1604 | 15 | US-10-938-842A-1034 |
| 7 | 20 | 1.9 | 1263 | 12 | US-10-938-842A-1034 |
| 8 | 19 | 1.8 | 34094 | 13 | US-09-938-842A-1034 |
| 9 | 19 | 1.8 | 34094 | 15 | US-10-938-842A-1034 |
| 10 | 19 | 1.8 | 34094 | 16 | US-10-938-842A-1034 |
| 11 | 18 | 1.7 | 277 | 9 | US-09-938-842A-1034 |
| 12 | 18 | 1.7 | 364 | 11 | US-09-938-842A-1034 |
| 13 | 18 | 1.7 | 377 | 11 | US-09-938-842A-1034 |
| 14 | 18 | 1.7 | 424 | 9 | US-09-938-842A-1034 |
| 15 | 18 | 1.7 | 424 | 15 | US-10-938-842A-1034 |

| | | | | | | |
|------|----|-----|--------|----|----------------------|-------------------|
| C 16 | 18 | 1.7 | 432 | 10 | US-09-983-965-5485 | Sequence 5485, Ap |
| C 17 | 18 | 1.7 | 475 | 9 | US-09-770-444-64 | Sequence 64, Appl |
| C 18 | 18 | 1.7 | 572 | 13 | US-10-027-632-216038 | Sequence 216038, |
| C 19 | 18 | 1.7 | 572 | 14 | US-10-027-632-216038 | Sequence 216038, |
| C 20 | 18 | 1.7 | 626 | 9 | US-09-770-149-739 | Sequence 739, App |
| C 21 | 18 | 1.7 | 1119 | 12 | US-10-369-493-32488 | Sequence 32488, A |
| C 22 | 18 | 1.7 | 1487 | 12 | US-10-062-674-1944 | Sequence 1944, Ap |
| C 23 | 18 | 1.7 | 2577 | 10 | US-09-938-842A-337 | Sequence 337, App |
| C 24 | 18 | 1.7 | 2577 | 12 | US-09-938-842A-337 | Sequence 337, App |
| C 25 | 18 | 1.7 | 3086 | 12 | US-10-369-493-27151 | Sequence 27151, A |
| C 26 | 18 | 1.7 | 3757 | 12 | US-10-262-445-59 | Sequence 59, Appl |
| C 27 | 18 | 1.7 | 4721 | 12 | US-10-262-445-57 | Sequence 57, Appl |
| C 28 | 18 | 1.7 | 14353 | 13 | US-10-311-455-2408 | Sequence 2408, Ap |
| C 29 | 18 | 1.7 | 162450 | 13 | US-10-126-704-1 | Sequence 1, Appli |
| C 30 | 18 | 1.7 | 162450 | 15 | US-10-071-179-1 | Sequence 1, Appli |
| C 31 | 17 | 1.6 | 344 | 9 | US-09-770-791-787 | Sequence 787, App |
| C 32 | 17 | 1.6 | 385 | 10 | US-09-960-352-5500 | Sequence 5500, Ap |
| C 33 | 17 | 1.6 | 396 | 10 | US-09-938-842A-370 | Sequence 370, App |
| C 34 | 17 | 1.6 | 396 | 12 | US-09-938-842A-370 | Sequence 370, App |
| C 35 | 17 | 1.6 | 402 | 10 | US-09-960-352-1265 | Sequence 1265, Ap |
| C 36 | 17 | 1.6 | 406 | 10 | US-09-960-352-273 | Sequence 273, App |
| C 37 | 17 | 1.6 | 435 | 11 | US-09-918-995-1982 | Sequence 1982, Ap |
| C 38 | 17 | 1.6 | 442 | 15 | US-10-178-213-328 | Sequence 328, App |
| C 39 | 17 | 1.6 | 490 | 11 | US-09-770-961-675 | Sequence 675, App |
| C 40 | 17 | 1.6 | 498 | 9 | US-09-244-694-116 | Sequence 116, App |
| C 41 | 17 | 1.6 | 543 | 13 | US-10-027-632-323443 | Sequence 323443, |
| C 42 | 17 | 1.6 | 543 | 13 | US-10-027-632-324847 | Sequence 324847, |
| C 43 | 17 | 1.6 | 543 | 14 | US-10-027-632-323443 | Sequence 323443, |
| C 44 | 17 | 1.6 | 543 | 14 | US-10-027-632-324847 | Sequence 324847, |
| C 45 | 17 | 1.6 | 563 | 10 | US-09-796-692-6933 | Sequence 6933, Ap |

ALIGNMENTS

RESULT 1

US-09-938-842A-1034
Sequence 1034, Application US/09938842A
Patent No: US20030160379A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
TITLE OF INVENTION: SAME, AND METHODS OF USE
FILE REFERENCE: SCRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 1034
LENGTH: 1071
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-1034

| | | | | |
|-----------------------|--------------|--|------------|--------------|
| Query Match | 100.0% | Score 1071, | DB 10, | Length 1071, |
| Best Local Similarity | 100.0% | Pred. No. 0, | | |
| Matches 1071, | Conservative | 0, | Mismatches | 0, |
| | | | Indels | 0, |
| | | | Gaps | 0, |
| QY | 1 | ATGGCGACAATTCAGAGCTTGAAGAAGTTGCGAGCAAGATCAAACTTAAGAGCGGTT | 60 | |
| Db | 1 | ATGGCGACAATTCAGAGCTTGAAGAAGTTGCGAGCAAGATCAAACTTAAGAGCGGTT | 60 | |
| QY | 61 | GATCTAACCATCATCAACGGCGTTCAGAAAGTTCGAAACTTCAAGACCTTCCAGTAAT | 120 | |
| Db | 61 | GATCTAACCATCATCAACGGCGTTCAGAAAGTTCGAAACTTCAAGACCTTCCAGTAAT | 120 | |

121 CCCACAGTCTCGAGCCCAAGCCGAGCCGCTGATGCCGTGTTTCAATGTCCTTTA 180
121 CCCACAGTCTCGAGCCCAAGCCGAGCCGCTGATGCCGTGTTTCAATGTCCTTTA 180
181 GCTCCACGCTTTCGACGAGCACCACCTTGAAGAGCTTCGACTAAGACCTTCACAG 240
181 GCTCCACGCTTTCGACGAGCACCACCTTGAAGAGCTTCGACTAAGACCTTCACAG 240
241 AAGGTTGAAGGAGGAGGAGGATACGATGCTGCCACGCTGCGGTAGGATTTT 300
241 AAGGTTGAAGGAGGAGGAGGATACGATGCTGCCACGCTGCGGTAGGATTTT 300
301 CAATTAATCGAGATAGTTCACAAATCCGACGAGCAACGCTGATTCGATGAGTAA 360
301 CAATTAATCGAGATAGTTCACAAATCCGACGAGCAACGCTGATTCGATGAGTAA 360
361 AACGCTGAGCCGCGGATATATAGCCGACGAGGAGGAGGAGGAGGAGGAGGAGGAG 420
361 AACGCTGAGCCGCGGATATATAGCCGACGAGGAGGAGGAGGAGGAGGAGGAGGAG 420
421 TCGGTTAAGCAACCTTAAATATCCGACGAGCAACGCTGATTCGATGAGTAA 480
421 TCGGTTAAGCAACCTTAAATATCCGACGAGCAACGCTGATTCGATGAGTAA 480
481 AATCTGATGAAGAGAAACGCTTAAAGCACTTCTAAGAGTATATAGACATACGAG 540
481 AATCTGATGAAGAGAAACGCTTAAAGCACTTCTAAGAGTATATAGACATACGAG 540
541 GCGGTTTCAAGCTTCTCGGTTTGTAGTCCAAATGCCAGCAACGCTGATTCGATGAG 600
541 GCGGTTTCAAGCTTCTCGGTTTGTAGTCCAAATGCCAGCAACGCTGATTCGATGAG 600
601 CAAGCTCTGCATCATCTGCTGCTCAGCACTTCTGCGGCAAGGAGTATCCGATG 660
601 CAAGCTCTGCATCATCTGCTGCTCAGCACTTCTGCGGCAAGGAGTATCCGATG 660
661 TGGGCTATTCCATCAACGCAATGATTCGACGCTGAGAGCTTCTTCTGATTCACAA 720
661 TGGGCTATTCCATCAACGCAATGATTCGACGCTGAGAGCTTCTTCTGATTCACAA 720
721 ATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
721 ATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
781 TCGTCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
781 TCGTCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
841 GTTGTTCGAGCAGCGGCTTGTATCGGTTTCAGAGCTTTCGAGGTTTCGAAATTA 900
841 GTTGTTCGAGCAGCGGCTTGTATCGGTTTCAGAGCTTTCGAGGTTTCGAAATTA 900
901 GCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
901 GCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
961 ATTGCAACCAACGACGACGCTGAGAGCTTCTCCCTAGAGATATACGAGAAACAA 1020
961 ATTGCAACCAACGACGACGCTGAGAGCTTCTCCCTAGAGATATACGAGAAACAA 1020
1021 GAGCTTACAGGTTTCATGACACACACGAGCTGATTCGAGCAACCTGTA 1071
1021 GAGCTTACAGGTTTCATGACACACACGAGCTGATTCGAGCAACCTGTA 1071

RESULT 2

US-09-938-842a-1034
; Sequence 1034, Application US/09938842A
; Publication No. US20040009476A9
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel

APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
TITLE OF INVENTION: SAME, AND METHODS OF USE
FILE REFERENCE: SRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 1034
LENGTH: 1071
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-1034

Query Match 100.0%; Score 1071; DB 12; Length 1071;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1071; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCGCAATTCAGAGCTTGAAGAGTTGCGAGCAAGATCAAACTCTAAGAGCGGTT 60
DB 1 ATGCGCAATTCAGAGCTTGAAGAGTTGCGAGCAAGATCAAACTCTAAGAGCGGTT 60
QY 61 GATCTAACCATCATCAACGCGCTCAGAACTTCAAGAGCTTTCCTCAAGTAAAT 120
DB 61 GATCTAACCATCATCAACGCGCTCAGAACTTCAAGAGCTTTCCTCAAGTAAAT 120
QY 121 CCGACAGTGTCTGAGCCCAAGCGGAGCCGCTGATGCCGTGTTTCAATGTCCTTTA 180
DB 121 CCGACAGTGTCTGAGCCCAAGCGGAGCCGCTGATGCCGTGTTTCAATGTCCTTTA 180
QY 181 GCTCCACGCTTTCGACGAGCACCACCTTGAAGAGCTTCGACTAAGACCTTCACAG 240
DB 181 GCTCCACGCTTTCGACGAGCACCACCTTGAAGAGCTTCGACTAAGACCTTCACAG 240
QY 241 AAGGTTGAAGGAGGAGGAGGATACGATGCTGCCACGCTGCTGCTGCTGCTGCTGCTG 300
DB 241 AAGGTTGAAGGAGGAGGAGGATACGATGCTGCCACGCTGCTGCTGCTGCTGCTGCTG 300
QY 301 CAATTAATCGAGATAGTTCACAAATGCCAGCAACGCTGATTCGATGAGTAA 360
DB 301 CAATTAATCGAGATAGTTCACAAATGCCAGCAACGCTGATTCGATGAGTAA 360
QY 361 AACGCTGAGCCGCGGATATATAGCCGACGAGGAGGAGGAGGAGGAGGAGGAGGAG 420
DB 361 AACGCTGAGCCGCGGATATATAGCCGACGAGGAGGAGGAGGAGGAGGAGGAGGAG 420
QY 421 TCGGTTAAGCAACCTTAAATATCCGACGAGCAACGCTGATTCGATGAGTAA 480
DB 421 TCGGTTAAGCAACCTTAAATATCCGACGAGCAACGCTGATTCGATGAGTAA 480
QY 481 AATCTGATGAAGAGAAACGCTTAAAGCACTTCTAAGAGTATATAGACATACGAG 540
DB 481 AATCTGATGAAGAGAAACGCTTAAAGCACTTCTAAGAGTATATAGACATACGAG 540
QY 541 GCGGTTTCAAGCTTCTCGGTTTGTAGTCCAAATGCCAGCAACGCTGATTCGATGAG 600
DB 541 GCGGTTTCAAGCTTCTCGGTTTGTAGTCCAAATGCCAGCAACGCTGATTCGATGAG 600
QY 601 CAAGCTCTGCATCATCTGCTGCTCAGCACTTCTGCGGCAAGGAGTATCCGATG 660
DB 601 CAAGCTCTGCATCATCTGCTGCTCAGCACTTCTGCGGCAAGGAGTATCCGATG 660
QY 661 TGGGCTATTCCATCAACGCAATGATTCGACGCTGAGAGCTTCTTCTGATTCACAA 720
DB 661 TGGGCTATTCCATCAACGCAATGATTCGACGCTGAGAGCTTCTTCTGATTCACAA 720
QY 721 ATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780

721 ATCGTGGTCCGTCAGTATAGCTTATAGCTTTTCCCGCGCGCTGCTTCGCG 780
781 TCGTCTTACGTCGCGCTGTTTCAACAGAGCTTCCAGATGGCTAGACACCTCCCTTTACAA 840
781 TCGTCTTACGTCGCGCTGTTTCAACAGAGCTTCCAGATGGCTAGACACCTCCCTTTACAA 840
841 GTTGTTCGAAGCAGCGCTTTGTATCCGTTTTCAGAGCTTACGGTTCGAATTTATCAAGA 900
841 GTTGTTCGAAGCAGCGCTTTGTATCCGTTTTCAGAGCTTACGGTTCGAATTTATCAAGA 900
901 CGGAGCTCGGTTATGGCTCCGAGCTCAAGCTCAGCGGTAAACAACCGGTAGTTTCATCGTCA 960
901 CGGAGCTCGGTTATGGCTCCGAGCTCAAGCTCAGCGGTAAACAACCGGTAGTTTCATCGTCA 960
961 ATTGCAACAACAACGACGACACCGCTAGAGACTTCTCCCTAGAGATATACGAGAAACA 1020
961 ATTGCAACAACAACGACGACACCGCTAGAGACTTCTCCCTAGAGATATACGAGAAACA 1020
1021 GAGCTTCCAGCTTATGATGAGCACCACACGAGCTCATCGAACCACTGA 1071
1021 GAGCTTCCAGCTTATGATGAGCACCACACGAGCTCATCGAACCACTGA 1071

RESULT 3

S-09-924-035A-502/c

Sequence 502, Application US/09924035A

Patent No. US20020142319A1

GENERAL INFORMATION:

APPLICANT: Grlach, Jiri

TITLE OF INVENTION: Expressed Sequences of Arabidopsis

FILE OF INVENTION: thaliana

CURRENT APPLICATION NUMBER: US/09/924,035A

CURRENT FILING DATE: 2000-08-11

PRIOR APPLICATION NUMBER: US 60/148,784

PRIOR FILING DATE: 1999-08-13

NUMBER OF SEQ ID NOS: 900

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 502

LENGTH: 460

TYPE: DNA

ORGANISM: Arabidopsis thaliana

FEATURE:

NAME/KEY: misc feature

LOCATION: (1)-(460)

OTHER INFORMATION: n = A,T,C or G

S-09-924-035A-502

Query Match 27.0%; Score 289; DB 10; Length 460;
Best Local Similarity 99.6%; Pred. No. 9.1e-147;
Matches 459; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
Y 600 GCAAGCTTGCATCATCCATCTGGCTCAGCAACTTCTGCGGCAAGGAATGTATCCGAT 659
b 460 GCAAGCTTGCATCATCCATCTGGCTCAGCAACTTCTGCGGCAAGGAATGTATCCGAT 401
660 GTGGCTATTCCATCAACGCAATGATTCGACGCTCGAGCTTCTTGTGATCCACA 719
b 400 GTGGCTATTCCATCAACGCAATGATTCGACGCTCGAGCTTCTTGTGATCCACA 341
720 AATCGTGGTCCGTCGATCAGCTCAGCTTATAGCTTTTCCCGCCCGCGCTGCTTCGCC 779
b 340 AATCGTGGTCCGTCGATCAGCTCAGCTTATAGCTTTTCCCGCCCGCGCGNG-TTCGCC 282
780 GTGCTTTACGTCGCGCTGTTCAACAGGCTTCCACGATGCTAGACACCTCCTTTACA 839
b 281 GTGCTTTACGTCGCGCTGTTCAACAGGCTTCCACGATGCTAGACACCTCCTTTACA 222
840 AGTTGTTCCAGCAGCGGCTTTGTATCCGTTTTCAGAGCTTAGCGGTTTCGAATTTATCAAG 899
b 221 AGTTGTTCCAGCAGCGGCTTTGTATCCGTTTTCAGAGCTTAGCGGTTTCGAATTTATCAAG 162

QY 900 AGGAGCTCGGTTATGGCTCCGAGCTCAAGCTCAGGCGTAAACAACCGGTAGTTTCATCGTC 959
Db 161 AGGAGCTCGGTTATGGCTCCGAGCTCAAGCTCAGGCGTAAACAACCGGTAGTTTCATCGTC 102
QY 960 AATTGCAACAACAACGAGCAGCAGCTGAGAGCTTCTCCCTAGAGATATACGAGAAACA 1019
Db 101 AATTGCAACAACAACGAGCAGCAGCTGAGAGCTTCTCCCTAGAGATATACGAGAAACA 42
QY 1020 AGAGCTTCCAGCTTATGATGAGCACCACCAACGACCGGTCTAT 1060
Db 41 AGAGCTTCCAGCTTATGATGAGCACCACCAACGACCGGTCTAT 1

RESULT 4

US-09-770-444-615/c

Sequence 615, Application US/09770444

Patent No. US20020023280A1

GENERAL INFORMATION:

APPLICANT: Grlach, Jiri

APPLICANT: An, Yong-Qiang

APPLICANT: Hamilton, Carol M.

APPLICANT: Price, Jennifer L.

APPLICANT: Raines, Tracy M.

APPLICANT: Yu, Yang

APPLICANT: Rameaka, Joshua G.

APPLICANT: Page, Amy

APPLICANT: Matthew, Abraham V.

APPLICANT: Ledford, Brooke L.

APPLICANT: Woessner, Jeffrey P.

APPLICANT: Haas, William David

APPLICANT: Garcia, Carlos A.

APPLICANT: Kriker, Maja

APPLICANT: Slader, Ted

APPLICANT: Davis, Keith R.

APPLICANT: Allen, Keith

APPLICANT: Hoffman, Neil

APPLICANT: Hurlan, Patrick

TITLE OF INVENTION: Expressed Sequences of Arabidopsis

FILE OF INVENTION: thaliana

CURRENT APPLICATION NUMBER: US/09/770,444

CURRENT FILING DATE: 2001-01-26

PRIOR APPLICATION NUMBER: 60/178,502

PRIOR FILING DATE: 2000-01-27

NUMBER OF SEQ ID NOS: 999

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 615

LENGTH: 453

TYPE: DNA

ORGANISM: Arabidopsis thaliana

FEATURE:

NAME/KEY: misc feature

LOCATION: (1)....(453)

OTHER INFORMATION: n = A,T,C or G

US-09-770-444-615

Query Match 26.7%; Score 286; DB 9; Length 453;
Best Local Similarity 100.0%; Pred. No. 3.9e-145;
Matches 286; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 775 TCGCGTGTCTTACGTGCGCGCTGTTCAACAGGCTTCCAGATGGCTAGACACCTCCT 834
Db 286 TCGCGTGTCTTACGTGCGCGCTGTTCAACAGGCTTCCAGATGGCTAGACACCTCCT 227
QY 835 TTACAAGTTGTTCCAGCAGCGGCTTTGTTATCCGTTTTCAGAGCTTAGCGGTTCAATTTA 894
Db 226 TTACAAGTTGTTCCAGCAGCGGCTTTGTTATCCGTTTTCAGAGCTTAGCGGTTCAATTTA 167
QY 895 TCAAGAGCAGCTCGGTTATGGCTCCGAGCTCAAGCTCAGCGGTAAACAACCGGTAGTTCA 954
Db 166 TCAAGAGCAGCTCGGTTATGGCTCCGAGCTCAAGCTCAGCGGTAAACAACCGGTAGTTCA 107
QY 955 TCGTCAATGCAACAACAACGACGCAACGCTGAGAGCTTCTCCCTAGAGATATACGAG 1014

Db 106 TCGTCAATTGCAACAACACGACGACACGCTGAGAGACTTCTCCCTAGAGATATACGAG 47
QY 1015 AAACAAGAGCTTCCAGGTTTCATGACGACACACACACGCGGTCAT 1060
Db 46 AAACAAGAGCTTCCAGGTTTCATGACGACACACACACGCGGTCAT 1

RESULT 5

US-09-770-696-257
; Sequence 257, Application US/09770696
; Patent No. US20010044940A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krickler, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: thaliana
; FILE REFERENCE: 2031US (PARA-020PRV)
; CURRENT APPLICATION NUMBER: US/09/770,696
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/178,278
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 911
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 257
; LENGTH: 185
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-696-257

Query Match 17.3%; Score 185; DB 9; Length 185;
Best Local Similarity 100.0%; Pred. No. 4.9e-90;
Matches 185; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 35 GCAAGATCAAACTTAAGAGCGTTGATCTAACCATCATCAACGGCGTCAGAACTCG 94
Db 1 GCAAGATCAAACTTAAGAGCGTTGATCTAACCATCATCAACGGCGTCAGAACTCG 60
QY 95 AAATTTCAAGACCTTTCCAAAGTAAATCCCAAGTGTCTCGAGCCCAAGGCGGACCGG 154
Db 61 AAATTTCAAGACCTTTCCAAAGTAAATCCCAAGTGTCTCGAGCCCAAGGCGGACCGG 120
QY 155 TGATGCGGTGTTTTCAATGCTTTAGCTCCAGGCTTCGACGAGACCACTTGAAGA 214
Db 121 TGATGCGGTGTTTTCAATGCTTTAGCTCCAGGCTTCGACGAGACCACTTGAAGA 180
QY 215 GAGCT 219
Db 181 GAGCT 185

RESULT 6

US-10-295-403-147
; Sequence 147, Application US/10295403
; Publication No. US20030101481A1
; GENERAL INFORMATION:

; APPLICANT: Heard, Jacqueline
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Adam, Luc
; APPLICANT: Broun, Pierre
; APPLICANT: Pineda, Omalta
; APPLICANT: Reuber, Lynne
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Keddle, James
; APPLICANT: Zhang, James
; APPLICANT: Benito, Maria-Ines
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Fromm, Mike
; TITLE OF INVENTION: PLANT GENE SEQUENCES I
; FILE REFERENCE: MBI-0003
; CURRENT APPLICATION NUMBER: US/10/295,403
; CURRENT FILING DATE: 2002-11-15
; PRIOR APPLICATION NUMBER: US/09/394,519
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: 60/101,349
; PRIOR FILING DATE: 1998-09-22
; PRIOR APPLICATION NUMBER: 60/103,312
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: 60/108,734
; PRIOR FILING DATE: 1998-11-17
; PRIOR APPLICATION NUMBER: 60/113,409
; PRIOR FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 147
; LENGTH: 1604
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (143)..(1345)
; OTHER INFORMATION: G802
US-10-295-403-147

Query Match 2.1%; Score 23; DB 15; Length 1604;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 TCGACTAAAGACCGTCACAGAA 242
Db 317 TCGACTAAAGACCGTCACAGAA 339

RESULT 7

US-10-369-493-37704/c
; Sequence 37704, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 37704
; LENGTH: 1263
; TYPE: DNA
; ORGANISM: Pseudomonas fluorescens
US-10-369-493-37704

Query Match 1.9%; Score 20; DB 12; Length 1263;
Best Local Similarity 100.0%; Pred. No. 5.3;

; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30316
; LENGTH: 364
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(364)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-30316

Query Match 1.7%; Score 18; DB 11; Length 364;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 603 AGCTCTGGCATCATCCAC 620
Db 218 AGCTCTGGCATCATCCAC 201

RESULT 13
US-09-918-995-34549/c
; Sequence 34549, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED

; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34549
; LENGTH: 377
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-34549

Query Match 1.7%; Score 18; DB 11; Length 377;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 603 AGCTCTGGCATCATCCAC 620
Db 226 AGCTCTGGCATCATCCAC 209

RESULT 14
US-09-777-564-1604/c
; Sequence 1604, Application US/09777564
; Patent No. US20020022591A1
; GENERAL INFORMATION:
; APPLICANT: Marion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.493
; CURRENT APPLICATION NUMBER: US/09/777,564
; CURRENT FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 1730
; SOFTWARE: FastSeq for Window Version 4.0
; SEQ ID NO 1604
; LENGTH: 424
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-777-564-1604

Query Match 1.7%; Score 18; DB 9; Length 424;

Best Local Similarity 100.0%; Pred. No. 63;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 603 AGCTCTGGCATCATCCAC 620
Db 83 AGCTCTGGCATCATCCAC 66

RESULT 15
US-10-015-219-1604/c
; Sequence 1604, Application US/10015219
; Publication No. US20030008299A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.493C1
; CURRENT APPLICATION NUMBER: US/10/015,219
; CURRENT FILING DATE: 2002-03-02
; NUMBER OF SEQ ID NOS: 1739
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1604
; LENGTH: 424
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-015-219-1604

Query Match 1.7%; Score 18; DB 15; Length 424;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 603 AGCTCTGGCATCATCCAC 620
Db 83 AGCTCTGGCATCATCCAC 66

Search completed: February 2, 2004, 08:33:14
Job time : 808.301 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

DM nucleic - nucleic search, using sw model
Run on: February 2, 2004, 01:11:01 ; Search time 6363.2 Seconds
(without alignments)
10781.591 Million cell updates/sec

Title: US-09-938-842a-3729

Perfect score: 1677

Sequence: 1 ggtaagcgttttactatg.....tttctcagctatatattta 1677

Scoring table: OLIGO NUC
Gapop_60.0 , Gapext 60.0

Searched: 2888711 seqs, 20454813386 residues

Word size: 0

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

GenEmbl:

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.in.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sv.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.mu.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pin.*
35: em.htg.rod.*
36: em.htg.nam.*
37: em.htg.vrt.*
38: em.sv.*
39: em.higo.hum.*
40: em.higo.mus.*
41: em.higo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB ID | Description |
|------------|-------|---------------|--------|-------|--------------------|
| 1 | 1677 | 100.0 | 1677 | 6 | AX509034 Sequence |
| 2 | 1677 | 100.0 | 91854 | 8 | AC003680 Arabidops |
| 3 | 193 | 11.5 | 1483 | 8 | AF085354 Arabidops |
| 4 | 102 | 6.1 | 1660 | 8 | AY056214 Arabidops |
| 5 | 24 | 1.4 | 250957 | 2 | AC106200 Rattus no |
| 6 | 24 | 1.4 | 282156 | 2 | AC119331 Rattus no |
| 7 | 23 | 1.4 | 101 | 6 | AX114617 Sequence |
| 8 | 23 | 1.4 | 101 | 6 | AX114618 Sequence |
| 9 | 23 | 1.4 | 101 | 6 | AX114743 Sequence |
| 10 | 23 | 1.4 | 101 | 6 | AX114744 Sequence |
| 11 | 23 | 1.4 | 101 | 6 | AX118672 Sequence |
| 12 | 23 | 1.4 | 101 | 6 | AX118673 Sequence |
| 13 | 23 | 1.4 | 101 | 6 | AX118673 Sequence |
| 14 | 23 | 1.4 | 101 | 6 | AX128089 Sequence |
| 15 | 23 | 1.4 | 101 | 6 | AX128090 Sequence |
| 16 | 23 | 1.4 | 101 | 6 | AX644690 Sequence |
| 17 | 23 | 1.4 | 7240 | 6 | AX114613 Sequence |
| 18 | 23 | 1.4 | 7240 | 6 | AX114739 Sequence |
| 19 | 23 | 1.4 | 7240 | 6 | AX118669 Sequence |
| 20 | 23 | 1.4 | 7240 | 6 | AX128086 Sequence |
| 21 | 23 | 1.4 | 7240 | 6 | AX644687 Sequence |
| 22 | 23 | 1.4 | 7240 | 9 | HUMINSRD |
| 23 | 23 | 1.4 | 7641 | 9 | M29929 Human insul |
| 24 | 23 | 1.4 | 47467 | 2 | AC135456 Rattus no |
| 25 | 23 | 1.4 | 72172 | 9 | AC010311 Homo sapi |
| 26 | 23 | 1.4 | 125227 | 2 | AC102324 Mus muscu |
| 27 | 23 | 1.4 | 140868 | 2 | AC142238 Rattus no |
| 28 | 23 | 1.4 | 174141 | 2 | AC102313 Mus muscu |
| 29 | 23 | 1.4 | 187783 | 2 | AC137462 Rattus no |
| 30 | 23 | 1.4 | 241280 | 2 | AC115158 Rattus no |
| 31 | 23 | 1.4 | 248948 | 2 | AC097421 Rattus no |
| 32 | 23 | 1.4 | 254082 | 2 | AC101655 Mus muscu |
| 33 | 23 | 1.4 | 260809 | 2 | AC126820 Rattus no |
| 34 | 23 | 1.4 | 267375 | 2 | AC095339 Rattus no |
| 35 | 22 | 1.3 | 47328 | 6 | AX059532 Sequence |
| 36 | 22 | 1.3 | 103632 | 8 | AF074021 Arabidops |
| 37 | 22 | 1.3 | 164679 | 9 | AC005921 Homo sapi |
| 38 | 22 | 1.3 | 179155 | 10 | AL772401 Mouse DNA |
| 39 | 22 | 1.3 | 189484 | 2 | BX470113 Danio rer |
| 40 | 22 | 1.3 | 200001 | 8 | AL61501 Arabidops |
| 41 | 22 | 1.3 | 216886 | 2 | AC127846 Rattus no |
| 42 | 22 | 1.3 | 236723 | 2 | AC106621 Rattus no |
| 43 | 21 | 1.3 | 330 | 6 | AX210412 Sequence |
| 44 | 21 | 1.3 | 394 | 6 | AX245777 Sequence |
| 45 | 21 | 1.3 | 407 | 6 | AX246703 Sequence |

ALIGNMENTS

RESULT 1
AX509034
LOCUS AX509034 1677 bp DNA linear PAT 27-SEP-2002
DEFINITION Sequence 3729 from Patent WO0216655.
ACCESSION AX509034
VERSION AX509034.1 GI:23390271
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi
1
REFERENCE
AUTHORS Harper J.F., Kreps, J., Wang, X. and Zhu, T.
TITLE Stress-regulated genes of plants, transgenic plants containing

same, and methods of use
Patent: WO 021655-A 3729 28-FEB-2002;
The Scripps Research Institute (US); Syngenta Participations AG
(CH)

FEATURES
source Location/Qualifiers
1..1677
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"

BASE COUNT 508 a 335 g 542 t
ORIGIN

Query Match 100.0%; Score 1677; DB 6; Length 1677;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1677; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTAAAGCGTTTACTTATGCTTATATGCAACGGAAGATATGCAATGTTGGAAATGC 60
DB 1 GGTAAAGCGTTTACTTATGCTTATATGCAACGGAAGATATGCAATGTTGGATGC 60

QY 61 TTTTTCAGATCATCAAGGCTCTACAGATTTCTTAGGGAATGGTTTCAGGCTTTGTTA 120
DB 61 TTTTTCAGATCATCAAGGCTCTACAGATTTCTTAGGGAATGGTTTCAGGCTTTGTTA 120

QY 121 GAAATGTGTTTATGCAACAGGTAGAGACATAACCATAGACAGATGTCTCGAAGAGA 180
DB 121 GAAATGTGTTTATGCAACAGGTAGAGACATAACCATAGACAGATGTCTCGAAGAGA 180

QY 181 TAAGCTTCTCTATGTCTAAAGAAATGACCGATACGAATTAATAAAGAT 240
DB 181 TAAGCTTCTCTATGTCTAAAGAAATGACCGATACGAATTAATAAAGAT 240

QY 241 TAAATGTTTGTAGAAATACTACATTTATGTAATGTAATGTTGTTAGTGAAGT 300
DB 241 TAAATGTTTGTAGAAATACTACATTTATGTAATGTAATGTTGTTAGTGAAGT 300

QY 301 AAAAACAATCGGAATCCAAACCTCAAAATTTACCAATCAGCCCAATTTATGTTGGG 360
DB 301 AAAAACAATCGGAATCCAAACCTCAAAATTTACCAATCAGCCCAATTTATGTTGGG 360

QY 361 TAATGAATGTATGCTGATGGTAGGCAAAAGTTGGTGGCTGGCAATTAACAATATC 420
DB 361 TAATGAATGTATGCTGATGGTAGGCAAAAGTTGGTGGCTGGCAATTAACAATATC 420

QY 421 CCTCTGTGTGGACCGAACTCTGTAATCGGAAAGTGGAAACCACTGGTTTAACTTTT 480
DB 421 CCTCTGTGTGGACCGAACTCTGTAATCGGAAAGTGGAAACCACTGGTTTAACTTTT 480

QY 481 AAGCTAAAGGTTTACACCGGTTTACCGGTTTAAATTTGGTGTGTTTAAATCTAATCCC 540
DB 481 AAGCTAAAGGTTTACACCGGTTTACCGGTTTAAATTTGGTGTGTTTAAATCTAATCCC 540

QY 541 GGATCCGTTTGTGTTTAAATCTCAAGCCACGTTATCGCCCAATTTGATTTTGGTGTG 600
DB 541 GGATCCGTTTGTGTTTAAATCTCAAGCCACGTTATCGCCCAATTTGATTTTGGTGTG 600

QY 601 GGTAGGAATGTTGGGTCGAATAGTTGGGCTAGCCCTCAACAAATGTGTGGAATGAAG 660
DB 601 GGTAGGAATGTTGGGTCGAATAGTTGGGCTAGCCCTCAACAAATGTGTGGAATGAAG 660

QY 661 AGATAGGTTCCAGCTCAGGCCACATTCATATTTGTTTGGTGTGTTTAAATCTAATCCC 720
DB 661 AGATAGGTTCCAGCTCAGGCCACATTCATATTTGTTTGGTGTGTTTAAATCTAATCCC 720

QY 721 TGCTTACGCTCCCTCTTTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 780
DB 721 TGCTTACGCTCCCTCTTTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 780

QY 781 CCGGAACAAGTACCAACGAATCAAAATAGTTTGAATCGGTTTACATCTAGTTACCGTCG 840
DB 781 CCGGAACAAGTACCAACGAATCAAAATAGTTTGAATCGGTTTACATCTAGTTACCGTCG 840

QY 841 AACTTACAATCATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 900

DB 841 AACTTACAATCATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 900
QY 901 CCGGATTTGTAACAGTACACAGTACATTAAGTATGCGTATATGATGACCGGTTTAAATC 960
DB 901 CCGGATTTGTAACAGTACACAGTACATTAAGTATGCGTATATGATGACCGGTTTAAATC 960

QY 961 AAAGGACGACGATAGGAGGATTTTGAATCTCTGGAAGAGGATTTTCCATAGACACTA 1020
DB 961 AAAGGACGACGATAGGAGGATTTTGAATCTCTGGAAGAGGATTTTCCATAGACACTA 1020

QY 1021 ATTAGCTTTTGGTGGCGAGCCTTGTGACCTACATTAATGGGGTCCACCCCAAGTATG 1080
DB 1021 ATTAGCTTTTGGTGGCGAGCCTTGTGACCTACATTAATGGGGTCCACCCCAAGTATG 1080

QY 1081 GGCCTTACAGCTTTTCCATTAATTAAGTAAATCTTTTTCCTTAAACCAATAAAAAT 1140
DB 1081 GGCCTTACAGCTTTTCCATTAATTAAGTAAATCTTTTTCCTTAAACCAATAAAAAT 1140

QY 1141 ATTGAAAATCTTTCCAAACCATAGAAAATTTGAATTTGATCAGCATGGAATTTTGTAC 1200
DB 1141 ATTGAAAATCTTTCCAAACCATAGAAAATTTGAATTTGATCAGCATGGAATTTTGTAC 1200

QY 1201 AAAGCTAGGTATTTTCTTTCGGAGTGTACTAGTAACTAGTAACTAACCAAGTGT 1260
DB 1201 AAAGCTAGGTATTTTCTTTCGGAGTGTACTAGTAACTAGTAACTAACCAAGTGT 1260

QY 1261 TTCTGATTTTGGATTTTGAAGCTTTTCTTAGGTTTAAATTTGATCAGCATGGAATTTTGTAC 1320
DB 1261 TTCTGATTTTGGATTTTGAAGCTTTTCTTAGGTTTAAATTTGATCAGCATGGAATTTTGTAC 1320

QY 1321 AAAAGAAAACATTTTGTGAAAAGAGAAATAAAGTTTACTTGGACCCCATTTGATCAGCATG 1380
DB 1321 AAAAGAAAACATTTTGTGAAAAGAGAAATAAAGTTTACTTGGACCCCATTTGATCAGCATG 1380

QY 1381 TCCCATTAATTAATCTGATAGAGATAGACCAATGGAAGTGTATTTTTCACGTGGTACAA 1440
DB 1381 TCCCATTAATTAATCTGATAGAGATAGACCAATGGAAGTGTATTTTTCACGTGGTACAA 1440

QY 1441 TCGGAATGTTTCTTAAAGCTCATCGAACATCAGACCGGTTGATTTTTCGGCATCAA 1500
DB 1441 TCGGAATGTTTCTTAAAGCTCATCGAACATCAGACCGGTTGATTTTTCGGCATCAA 1500

QY 1501 AAAGCGTTTGAATACTATTCTCAGCTGTTTCTTCTGCTTATATATATATATCTCGACGATC 1560
DB 1501 AAAGCGTTTGAATACTATTCTCAGCTGTTTCTTCTGCTTATATATATATATCTCGACGATC 1560

QY 1561 ACATTTAGTAAATCTCTTGGACCGTGAACCGGTTAAACGATTTCTTCCATTTGATCC 1620
DB 1561 ACATTTAGTAAATCTCTTGGACCGTGAACCGGTTAAACGATTTCTTCCATTTGATCC 1620

QY 1621 GCTTTTAAACCACTCTGTCGTCATCTCCACCGTCCGTTTCTCTCAGCTATATTTTA 1677
DB 1621 GCTTTTAAACCACTCTGTCGTCATCTCCACCGTCCGTTTCTCTCAGCTATATTTTA 1677

RESULT 2
AC003680/c
LOCUS
DEFINITION Arabidopsis thaliana chromosome 2 BAC F17K2 genomic sequence,
complete sequence.
ACCESSION AC003680
VERSION AC003680.3 GI:20197048
KEYWORDS HTG.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis;
1 (bases 1 to 91854)
Lin, X., Kaul, S., Town, C.D., Benito, M.-I., Creasy, T.H., Haas, B.J.,
Wu, D., Maiti, R., Ronning, C.M., Koo, H., Fujii, C.Y., Utterback, T.R.,
Barnstead, M.E., Bowman, C.L., White, O., Niernman, W.C. and Fraser, C.M.

Arabidopsis thaliana chromosome 2 BAC F17K2 genomic sequence

Unpublished
2 (bases 1 to 91854)

Lin, X.

Direct Submission

Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA

3 (bases 1 to 91854)

Town, C.D. and Kaul, S.

Direct Submission

Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA. cdowen@tigr.org
On Apr 18, 2002 this sequence version replaced gi:6598396.
Address all correspondence to: at@tigr.org

BAC clone F17K2 is from Arabidopsis thaliana chromosome 2 and is
near the molecular marker(s) F1S.
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.

Genes were identified by a combination of several methods: Gene
prediction programs including Genscan+ (Chris Burge,
<http://CCR-08.mit.edu/GENSCAN.html>), GeneMarkHM (Mark Borodovsky,
<http://genemark.biology.gatech.edu/GeneMark/>), GlimmerA (a variant
of GlimmerW, see Mihaela Pertea,
<http://www.tigr.org/softlab/glimmer.htm>), and
GeneSplicer (Mihaela Pertea and Steven Salzberg, contact
mpertea@tigr.org), searches of the complete sequence against a
peptide database and the plant EST database at TIGR
(<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to
indicate the level of evidence for their annotation. Genes with
similarity to other proteins are named after the database hits.
Genes without significant peptide similarity but with EST
similarity are named as unknown proteins. Genes without protein
or EST similarity, that are predicted by more than two gene
prediction programs over most of their length are annotated as
hypothetical proteins. Genes encoding tRNAs are predicted by
tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>).
Simple repeats are identified by repeatmasker (Arian Smit,
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

Location/Qualifiers
1. 91854
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="2"
/map="C1C02E07"
/clone="F17K2"
complement(1..1281)
/note="overlap with BAC clone F4L23 (AC002387:1..1281)."
57..747
/gene="At2g45480"
/note="synonym: F17K2.27"
join(<57..242,335..>747)
/gene="At2g45480"
join(57..242,335..747)
/gene="At2g45480"
/note="Protein sequence is in conflict with the conceptual
translation"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAF18607.2"
/db_xref="GI:297953"
/translation="DESKPSVQKFFPEVSDKCLBAKFNRRKNDIARGRWNNMV
NGGLFHGFHSPDVLQERGCFLQGVETNEPGRCCRTQKWCSCDVLGSKYCD
KMERGMKKHPDVTNENHAGFSPLTVAVRVVPCKDGDQKHSVSMGTLPLR
VSDKSTSSCSTDTITDIALRGDDDEYLSPGV"
complement(1002..2388)
/gene="At2g45490"
/note="synonym: F17K2.2; contains a protein kinase domain
profile (PDOC00100)"
complement(join(<1002..1112,1208..1280,1366..1580,
1655..1750,1886..2128,2260..>2388))

FEATURES

source

1. 91854
/gene="At2g45510"
/note="synonym: F17K2.4"
join(<6429..6950,7026..7310,7386..7724,7806..8006,
8100..>8288)
/gene="At2g45510"
join(6429..6950,7026..7310,7386..7724,7806..8006,
8100..8288)
/gene="At2g45510"
/codon_start=1
/product="putative cytochrome P450"
/protein_id="AAC06153.1"
/db_xref="GI:2979544"

CDS

1. 91854
/gene="At2g45510"
/note="synonym: F17K2.4"
join(<6429..6950,7026..7310,7386..7724,7806..8006,
8100..>8288)
/gene="At2g45510"
join(6429..6950,7026..7310,7386..7724,7806..8006,
8100..8288)
/gene="At2g45510"
/codon_start=1
/product="putative cytochrome P450"
/protein_id="AAC06153.1"
/db_xref="GI:2979544"

misc_feature

1. 91854
/gene="At2g45510"
/note="synonym: F17K2.4"
join(<6429..6950,7026..7310,7386..7724,7806..8006,
8100..>8288)
/gene="At2g45510"
join(6429..6950,7026..7310,7386..7724,7806..8006,
8100..8288)
/gene="At2g45510"
/codon_start=1
/product="putative cytochrome P450"
/protein_id="AAC06153.1"
/db_xref="GI:2979544"

gene

1. 91854
/gene="At2g45510"
/note="synonym: F17K2.4"
join(<6429..6950,7026..7310,7386..7724,7806..8006,
8100..>8288)
/gene="At2g45510"
join(6429..6950,7026..7310,7386..7724,7806..8006,
8100..8288)
/gene="At2g45510"
/codon_start=1
/product="putative cytochrome P450"
/protein_id="AAC06153.1"
/db_xref="GI:2979544"

gene

1. 91854
/gene="At2g45510"
/note="synonym: F17K2.4"
join(<6429..6950,7026..7310,7386..7724,7806..8006,
8100..>8288)
/gene="At2g45510"
join(6429..6950,7026..7310,7386..7724,7806..8006,
8100..8288)
/gene="At2g45510"
/codon_start=1
/product="putative cytochrome P450"
/protein_id="AAC06153.1"
/db_xref="GI:2979544"

gene

1. 91854
/gene="At2g45510"
/note="synonym: F17K2.4"
join(<6429..6950,7026..7310,7386..7724,7806..8006,
8100..>8288)
/gene="At2g45510"
join(6429..6950,7026..7310,7386..7724,7806..8006,
8100..8288)
/gene="At2g45510"
/codon_start=1
/product="putative cytochrome P450"
/protein_id="AAC06153.1"
/db_xref="GI:2979544"

gene

1. 91854
/gene="At2g45510"
/note="synonym: F17K2.4"
join(<6429..6950,7026..7310,7386..7724,7806..8006,
8100..>8288)
/gene="At2g45510"
join(6429..6950,7026..7310,7386..7724,7806..8006,
8100..8288)
/gene="At2g45510"
/codon_start=1
/product="putative cytochrome P450"
/protein_id="AAC06153.1"
/db_xref="GI:2979544"

gene

1. 91854
/gene="At2g45510"
/note="synonym: F17K2.4"
join(<6429..6950,7026..7310,7386..7724,7806..8006,
8100..>8288)
/gene="At2g45510"
join(6429..6950,7026..7310,7386..7724,7806..8006,
8100..8288)
/gene="At2g45510"
/codon_start=1
/product="putative cytochrome P450"
/protein_id="AAC06153.1"
/db_xref="GI:2979544"

gene

1. 91854
/gene="At2g45510"
/note="synonym: F17K2.4"
join(<6429..6950,7026..7310,7386..7724,7806..8006,
8100..>8288)
/gene="At2g45510"
join(6429..6950,7026..7310,7386..7724,7806..8006,
8100..8288)
/gene="At2g45510"
/codon_start=1
/product="putative cytochrome P450"
/protein_id="AAC06153.1"
/db_xref="GI:2979544"

gene

1. 91854
/gene="At2g45510"
/note="synonym: F17K2.4"
join(<6429..6950,7026..7310,7386..7724,7806..8006,
8100..>8288)
/gene="At2g45510"
join(6429..6950,7026..7310,7386..7724,7806..8006,
8100..8288)
/gene="At2g45510"
/codon_start=1
/product="putative cytochrome P450"
/protein_id="AAC06153.1"
/db_xref="GI:2979544"

gene

1. 91854
/gene="At2g45510"
/note="synonym: F17K2.4"
join(<6429..6950,7026..7310,7386..7724,7806..8006,
8100..>8288)
/gene="At2g45510"
join(6429..6950,7026..7310,7386..7724,7806..8006,
8100..8288)
/gene="At2g45510"
/codon_start=1
/product="putative cytochrome P450"
/protein_id="AAC06153.1"
/db_xref="GI:2979544"

gene

1. 91854
/gene="At2g45510"
/note="synonym: F17K2.4"
join(<6429..6950,7026..7310,7386..7724,7806..8006,
8100..>8288)
/gene="At2g45510"
join(6429..6950,7026..7310,7386..7724,7806..8006,
8100..8288)
/gene="At2g45510"
/codon_start=1
/product="putative cytochrome P450"
/protein_id="AAC06153.1"
/db_xref="GI:2979544"

gene

1. 91854
/gene="At2g45510"
/note="synonym: F17K2.4"
join(<6429..6950,7026..7310,7386..7724,7806..8006,
8100..>8288)
/gene="At2g45510"
join(6429..6950,7026..7310,7386..7724,7806..8006,
8100..8288)
/gene="At2g45510"
/codon_start=1
/product="putative cytochrome P450"
/protein_id="AAC06153.1"
/db_xref="GI:2979544"

gene

1. 91854
/gene="At2g45510"
/note="synonym: F17K2.4"
join(<6429..6950,7026..7310,7386..7724,7806..8006,
8100..>8288)
/gene="At2g45510"
join(6429..6950,7026..7310,7386..7724,7806..8006,
8100..8288)
/gene="At2g45510"
/codon_start=1
/product="putative cytochrome P450"
/protein_id="AAC06153.1"
/db_xref="GI:2979544"

gene

1. 91854
/gene="At2g45510"
/note="synonym: F17K2.4"
join(<6429..6950,7026..7310,7386..7724,7806..8006,
8100..>8288)
/gene="At2g45510"
join(6429..6950,7026..7310,7386..7724,7806..8006,
8100..8288)
/gene="At2g45510"
/codon_start=1
/product="putative cytochrome P450"
/protein_id="AAC06153.1"
/db_xref="GI:2979544"

gene

1. 91854
/gene="At2g45510"
/note="synonym: F17K2.4"
join(<6429..6950,7026..7310,7386..7724,7806..8006,
8100..>8288)
/gene="At2g45510"
join(6429..6950,7026..7310,7386..7724,7806..8006,
8100..8288)
/gene="At2g45510"
/codon_start=1
/product="putative cytochrome P450"
/protein_id="AAC06153.1"
/db_xref="GI:2979544"

gene

1. 91854
/gene="At2g45510"
/note="synonym: F17K2.4"
join(<6429..6950,7026..7310,7386..7724,7806..8006,
8100..>8288)
/gene="At2g45510"
join(6429..6950,7026..7310,7386..7724,7806..8006,
8100..8288)
/gene="At2g45510"
/codon_start=1
/product="putative cytochrome P450"
/protein_id="AAC06153.1"
/db_xref="GI:2979544"

gene

1. 91854
/gene="At2g45510"
/note="synonym: F17K2.4"
join(<6429..6950,7026..7310,7386..7724,7806..8006,
8100..>8288)
/gene="At2g45510"
join(6429..6950,7026..7310,7386..7724,7806..8006,
8100..8288)
/gene="At2g45510"
/codon_start=1
/product="putative cytochrome P450"
/protein_id="AAC06153.1"
/db_xref="GI:2979544"

CDS

1. 91854
/gene="At2g45490"
complement(join(1002..1112,1208..1280,1366..1580,
1655..1750,1886..2128,2260..>2388))
/gene="At2g45490"
/codon_start=1
/product="putative protein kinase"
/protein_id="AAC06151.1"
/db_xref="GI:2979542"

CDS

1. 91854
/gene="At2g45490"
complement(join(1002..1112,1208..1280,1366..1580,
1655..1750,1886..2128,2260..>2388))
/gene="At2g45490"
/codon_start=1
/product="putative protein kinase"
/protein_id="AAC06151.1"
/db_xref="GI:2979542"

CDS

1. 91854
/gene="At2g45490"
complement(join(1002..1112,1208..1280,1366..1580,
1655..1750,1886..2128,2260..>2388))
/gene="At2g45490"
/codon_start=1
/product="putative protein kinase"
/protein_id="AAC06151.1"
/db_xref="GI:2979542"

CDS

1. 91854
/gene="At2g45490"
complement(join(1002..1112,1208..1280,1366..1580,
1655..1750,1886..2128,2260..>2388))
/gene="At2g45490"
/codon_start=1
/product="putative protein kinase"
/protein_id="AAC06151.1"
/db_xref="GI:2979542"

CDS

1. 91854
/gene="At2g45490"
complement(join(1002..1112,1208..1280,1366..1580,
1655..1750,1886..2128,2260..>2388))
/gene="At2g45490"
/codon_start=1
/product="putative protein kinase"
/protein_id="AAC06151.1"
/db_xref="GI:2979542"

CDS

1. 91854
/gene="At2g45490"
complement(join(1002..1112,1208..1280,1366..1580,
1655..1750,1886..2128,2260..>2388))
/gene="At2g45490"
/codon_start=1
/product="putative protein kinase"
/protein_id="AAC06151.1"
/db_xref="GI:2979542"

CDS

1. 91854
/gene="At2g45490"
complement(join(1002..1112,1208..1280,1366..1580,
1655..1750,1886..2128,2260..>2388))
/gene="At2g45490"
/codon_start=1
/product="putative protein kinase"
/protein_id="AAC06151.1"
/db_xref="GI:2979542"

CDS

1. 91854
/gene="At2g45490"
complement(join(1002..1112,1208..1280,1366..1580,
1655..1750,1886..2128,2260..>2388))
/gene="At2g45490"
/codon_start=1
/product="putative protein kinase"
/protein_id="AAC06151.1"
/db_xref="GI:2979542"

CDS

1. 91854
/gene="At2g45490"
complement(join(1002..1112,1208..1280,1366..1580,
1655..1750,1886..2128,2260..>2388))
/gene="At2g45490"
/codon_start=1
/product="putative protein kinase"
/protein_id="AAC06151.1"
/db_xref="GI:2979542"

CDS

1. 91854
/gene="At2g45490"
complement(join(1002..1112,1208..1280,1366..1580,
1655..1750,1886..2128,2260..>2388))
/gene="At2g45490"
/codon_start=1
/product="putative protein kinase"
/protein_id="AAC06151.1"
/db_xref="GI:2979542"

CDS

1. 91854
/gene="At2g45490"
complement(join(1002..1112,1208..1280,1366..1580,
1655..1750,1886..2128,2260..>2388))
/gene="At2g45490"
/codon_start=1
/product="putative protein kinase"
/protein_id="AAC06151.1"
/db_xref="GI:2979542"

CDS

1. 91854
/gene="At2g45490"
complement(join(1002..1112,1208..1280,1366..1580,
1655..1750,1886..2128,2260..>2388))
/gene="At2g45490"
/codon_start=1
/product="putative protein kinase"
/protein_id="AAC06151.1"
/db_xref="GI:2979542"

CDS

1. 91854
/gene="At2g45490"
complement(join(1002..1112,1208..1280,1366..1580,
1655..1750,1886..2128,2260..>2388))
/gene="At2g45490"
/codon_start=1
/product="putative protein kinase"
/protein_id="AAC06151.1"
/db_xref="GI:2979542"

CDS

1. 91854
/gene="At2g45490"
complement(join(1002..1112,1208..1280,1366..1580,
1655..1750,1886..2128,2260..>2388))
/gene="At2g45490"
/codon_start=1
/product="putative protein kinase"
/protein_id="AAC06151.1"
/db_xref="GI:2979542"

CDS

1. 91854
/gene="At2g45490"
complement(join(1002..1112,1208..1280,1366..1580,
1655..1750,1886..2128,2260..>2388))
/gene="At2g45490"
/codon_start=1
/product="putative protein kinase"
/protein_id="AAC06151.1"
/db_xref="GI:2979542"

CDS

1. 91854
/gene="At2g45490"
complement(join(1002..1112,1208..1280,1366..1580,
1655..1750,1886..2128,2260..>2388))
/gene="At2g45490"
/codon_start=1
/product="putative protein kinase"
/protein_id="AAC06151.1"
/db_xref="GI:2979542"

CDS

1. 91854
/gene="At2g45490"
complement(join(1002..1112,1208..1280,1366..1580,
1655..1750,1886..2128,2260..>2388))
/gene="At2g45490"
/codon_start=1
/product="putative protein kinase"
/protein_id="AAC06151.1"
/db_xref="GI:2979542"

CDS

1. 91854
/gene="At2g45490"
complement(join(1002..1112,1208..1280,1366..1580,
1655..1750,1886..2128,2260..>2388))
/gene="At2g45490"
/codon_start=1
/product="putative protein kinase"
/protein_id="AAC06151.1"
/db_xref="GI:2979542"

CDS

1. 91854
/gene="At2g45490"
complement(join(1002..1112,1208..1280,1366..1580,
1655..1750,1886..2128,2260..>2388))
/gene="At2g45490"
/codon_start=1
/product="putative protein kinase"
/protein_id="AAC06151.1"
/db_xref="GI:2979542"

CDS

1. 91854
/gene="At2g45490"
complement(join(1002..1112,1208..1280,1366..1580,
1655..1750,1886..2128,2260..>2388))
/gene="At2g45490"
/codon_start=1
/product="putative protein kinase"
/protein_id="AAC06151.1"
/db_xref="GI:2979542"

CDS

1. 91854
/gene="At2g4549

```
/protein_id="AAC06155.1"
/db_xref="GI:2979546"
/translation="MGGKGRKREKNVLAHHGGPARLPPPPDRSKODDVPSTLRILMN
YTPSPDSTKOVVEKKELKAEVVVPAATESDSDSVVEKKKKRKNQWDLREN
ELAIIDGRSKRKKYKWEAKQKKXKGTEDTLRENPKHQIRFGDVVQAPLKIA
VWPKARSTLSASGERLRLQADAIRSRKGTARPGVPIFAVMWQ"
9408..9427
repeat_region
/rpt_family=" (CA)n"
complement(9578..9632)
/rpt_family=" (GA)n"
9822..10798
gene
/gene="At2g45530"
/notes="synonym: F17K2.6; supported by full length cDNA:
Ceres:35578"
9822..10798
mRNA
/gene="At2g45530"
9948..10670
CDS
/gene="At2g45530"
/codon_start=1
/product="expressed protein"
/protein_id="AAC06154.1"
/db_xref="GI:2979545"
/translation="MQSSSSSGDDLEKQOQOQKDKPOKOSANESNHLTIIVVCN
GDSRELVGQIPPEKEVSVNRGSSHEOCVCLQDKEVLRLGQCCRGGLAKAHS
CIDAPRTKSNOCIELQVAVNVTPETQPTNNVWRIDSRQEERERCGPFLW
VARSILGLMLVLSITLGVSAIPVNIIGVIVGLGTALRLTLEFCIEWSLRRA
VHRAVQRAEANNFSNTAIPPAL"
9988..10007
repeat_region
/rpt_family=" (CAG)n"
complement(10833..25573)
/gene="At2g45540"
/note="synonym: F17K2.7"
complement(join<10833..11272,11438..11560,11636..12378,
12454..12937,13023..13119,13254..13374,13464..13548,
13645..13794,13957..14073,14178..14264,14350..14418,
14520..14576,14730..14813,14926..14995,15396..15538,
15768..16244,16322..16465,16545..17093,17237..17431,
17612..17696,18015..18613,18944..19249,19633..19714,
19972..20048,20151..20264,20429..20533,20796..20930,
Query Match 100.0%; Score 1677; DB 8; Length 91854;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1677; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGTAAAGCGTTTACTTATGTTTATATGCAACGAGAAATATTGCCATTGTTGAATGC 60
DB 76628 GGTAAAGCGTTTACTTATGTTTATATGCAACGAGAAATATTGCCATTGTTGAATGC 76749
QY 61 TTTTTCAGATCATCAAGGCTCTACAGATTTCTTAGGGAATGTTTCAGGCTTTTGTTA 120
DB 76748 TTTTTCAGATCATCAAGGCTCTACAGATTTCTTAGGGAATGTTTCAGGCTTTTGTTA 76689
QY 121 GAAATTCGTTTATGCAACAGGTAGACATCAACATAGACAGATGTTCTGAGACA 180
DB 76588 GAAATTCGTTTATGCAACAGGTAGACATCAACATAGACAGATGTTCTGAGACA 76629
QY 181 TAAGCTTCTCTATGCTTAAAGAAATGACCGATACGAATAAACAAGCATCAATTAAGAT 240
DB 76628 TAAGCTTCTCTATGCTTAAAGAAATGACCGATACGAATAAACAAGCATCAATTAAGAT 76569
QY 241 TAAATGTTTGAAGAATACTACATTTATGTAAGAAATGTTGTTGTTAGTGAAGT 300
DB 76568 TAAATGTTTGAAGAATACTACATTTATGTAAGAAATGTTGTTGTTAGTGAAGT 76509
QY 301 AAAAATCATCGAATCCAAACCTCAAAATTTACCAATCAGCCCAAAATTTATGTCGTGGCG 360
DB 76508 AAAAATCATCGAATCCAAACCTCAAAATTTACCAATCAGCCCAAAATTTATGTCGTGGCG 76449
QY 361 TAATGAATGTTATGCTGATGGTAGGCAAAAGTTGGTGGCGAAATTAACAATATC 420
DB 76448 TAATGAATGTTATGCTGATGGTAGGCAAAAGTTGGTGGCGAAATTAACAATATC 76369
QY 421 CCTCTGTGGTGGACCGCAATCTGTAAATCGGAAGAGTGGNACCCACTTGGTTTAACTTTT 480
```

```
DB 76388 CCTCTGTGTGGACCGCAATCTGTAATCGGAAGGTGGAAACCACTTGGTTAACTTT 76329
QY 481 AAGCCTAAAAAGTTACTACCGGTTTGAACGGTTTATAATTTGGTTTAAATCTTAATCCC 540
DB 76328 AAGCCTAAAAAGTTACTACCGGTTTGAACGGTTTATAATTTGGTTTAAATCTTAATCCC 76269
QY 541 GGATCCGTTTGGTTTGAATCTCAAGCCACAGTTATCCCAATATTTTGATTTTGAAGTG 600
DB 76268 GGATCCGTTTGGTTTGAATCTCAAGCCACAGTTATCCCAATATTTTGATTTTGAAGTG 76209
QY 601 GGTAGGGAATGGTGGGTCGAATAGTTGGGCTAGCCCTCAACAATGTGTGGAACCTGAAG 660
DB 76208 GGTAGGGAATGGTGGGTCGAATAGTTGGGCTAGCCCTCAACAATGTGTGGAACCTGAAG 76149
QY 661 AGATAGGAGTCCAGCTCAGGCCACATTTCAATTTTCGTTTGTAGCTTTCTTTCTGG 720
DB 76148 AGATAGGAGTCCAGCTCAGGCCACATTTCAATTTTCGTTTGTAGCTTTCTTTCTGG 76089
QY 721 TGCTTACGGTCCCTCTTTCTGGTCCGTCGTATGTACAAGTAGCATAGCTAGTGGTTCAAA 780
DB 76088 TGCTTACGGTCCCTCTTTCTGGTCCGTCGTATGTACAAGTAGCATAGCTAGTGGTTCAAA 76029
QY 781 CCGAACAAGTACCAAGCAATCAAAATAGTTTGAATCGTTACATCTAGTTACCGTCG 840
DB 76028 CCGAACAAGTACCAAGCAATCAAAATAGTTTGAATCGTTACATCTAGTTACCGTCG 75969
QY 841 AACTTACAATCATTTCCGATTTTGAATCTGATTTCTAGTTCCGTTTGTATGTTTAATAT 900
DB 75968 AACTTACAATCATTTCCGATTTTGAATCTGATTTCTAGTTCCGTTTGTATGTTTAATAT 75909
QY 901 CCGATTTGTAAGTACCAAGTACCAAGTATCGGTATGCTATGATGTGACCGGTTTAAATC 960
DB 75908 CCGATTTGTAAGTACCAAGTACCAAGTATCGGTATGCTATGATGTGACCGGTTTAAATC 75849
QY 961 AAAGGACGAGCAGTAGGAGATTTTGAATCTCGGAAAGAGGATTTTCCATAGACACTA 1020
DB 75848 AAAGGACGAGCAGTAGGAGATTTTGAATCTCGGAAAGAGGATTTTCCATAGACACTA 75789
QY 1021 ATTAGCTTTTGTGGGCGAGCTTGTGACCTACATTAATGCGGTCCAAACCAAGTATG 1080
DB 75788 ATTAGCTTTTGTGGGCGAGCTTGTGACCTACATTAATGCGGTCCAAACCAAGTATG 75729
QY 1081 GCGTTACAGCTTTTTCATATAAATTAAGTAATCTTTTTCGCTAAACCAATTAAT 1140
DB 75728 GCGTTACAGCTTTTTCATATAAATTAAGTAATCTTTTTCGCTAAACCAATTAAT 75669
QY 1141 ATTGAAATCTTTCCAAACCATAGAAAAGTTAAATTTGATCAGCGATGGAATTTTGTAC 1200
DB 75668 ATTGAAATCTTTCCAAACCATAGAAAAGTTAAATTTGATCAGCGATGGAATTTTGTAC 75609
QY 1201 AAAGCTAGGTATTTTCATTTGGGAGTGTACTAGTAACTAGTAACTAGTAACTAGTAACT 1260
DB 75608 AAAGCTAGGTATTTTCATTTGGGAGTGTACTAGTAACTAGTAACTAGTAACTAGTAACT 75549
QY 1261 TTCTGATTTTGGATTTTGAAGCTTTTCTTAGGTTTAAATAAACAAGTATATTCTAAACAAT 1320
DB 75548 TTCTGATTTTGGATTTTGAAGCTTTTCTTAGGTTTAAATAAACAAGTATATTCTAAACAAT 75489
QY 1321 AAAAGAAAAACATTTTGTGAAAAGAGAAATTAAGTTTACTTGAACCCCTTGTACAGATGG 1380
DB 75488 AAAAGAAAAACATTTTGTGAAAAGAGAAATTAAGTTTACTTGAACCCCTTGTACAGATGG 75429
QY 1381 TCCCATATAATCTGATAGAGATAGAGCAATGGAAGAGTATTTGTTCCAGCTGGTACAA 1440
DB 75428 TCCCATATAATCTGATAGAGATAGAGCAATGGAAGAGTATTTGTTCCAGCTGGTACAA 75369
QY 1441 TCGAATAGGTCTTTTAAAGCTCATCGAAACATCATAGAGCGTTTGTATTTTCCGCACTAA 1500
DB 75368 TCGAATAGGTCTTTTAAAGCTCATCGAAACATCATAGAGCGTTTGTATTTTCCGCACTAA 75309
QY 1501 AAAGCGTTGATATACTATCTCACTTTTCTGCTTCTTATATATATATATATATATATATAT 1560
```

```
Db 75308 AAAGCGTTGTAATACTCTCACTGTTTCTTCCTGCTCCTATATATATATATATCTGACAGTTC 75249
2Y 1561 ACATTTAGTAATCTCTTGGAGCTGTAACGCCGTTAAACGATTTCTTTCCCATTTGATCC 1620
Db 75248 ACATTTAGTAATCTCTTGGAGCTGTAACGCCGTTAAACGATTTCTTTCCCATTTGATCC 75189
2Y 1621 GCATTTAAACAACTCTCGTCTGATCTATCCACCGTCCGTTTCTCTGAGCTATATTTTA 1677
Db 75188 GCATTTAAACAACTCTCGTCTGATCTATCCACCGTCCGTTTCTCTGAGCTATATTTTA 75132

RESULT 3
LOCUS AF085354 1483 bp mRNA linear PLN 18-MAY-1999
DEFINITION Arabidopsis thaliana shrunken seed protein (SSE1) mRNA, complete
cgs.
ACCESSION AF085354
VERSION AF085354.1 GI:4837732
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 1483)
AUTHORS Lin, Y., Sun, L., Nguyen, L.V., Rachubinski, R.A. and Goodman, H.M.
TITLE The Pex16p homolog SSE1 and storage organelle formation in
Arabidopsis seeds
JOURNAL Science 284 (5412), 328-330 (1999)
MEDLINE 99212146
PUBMED 10195899
FEATURES
SOURCE 1 (bases 1 to 1483)
AUTHORS Lin, Y., Sun, L., Nguyen, L.V. and Goodman, H.M.
TITLE Direct Submission
JOURNAL Submitted (19-AUG-1998) Molecular Biology, Massachusetts General
Hospital, 50 Blossom Street, Boston, MA 02114, USA
FEATURES
LOCATION/Qualifiers
1. .1483
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/cultivar="C24"
/db_xref="taxon:3702"
/chromosome="2"
/maps="near ve019"
1. .1483
/gene="SSE1"
122. .1225
/gene="SSE1"
/function="required for seed maturation"
/codon_start=1
/product="shrunken seed protein"
/protein_id="AAD30661.1"
/db_xref="GI:4837733"
/translation="MEAYKQWNRNREYVQSGFPANGLTWLLEKFSASEIGPEAVT
AFGLFTTINEHINATPFQHGSSNDPSLPFLIALKOLETVVEAAEHFYG
DKWNYITLAEAMKVALFRNSGYKLLQGGTETNEEKDSQNSQNRAGNSGN
LPGHGLQHNHPNWLGRAMSLSPSQNARTTSTPGWSRRIHQQAIVPPI
KERRTMSLELTKEGNGALFAIGELVITPLIVLPIRYGVRSWIPWALSVD
LQMGILLANSKWKGEKSKVHPSGPEKDLRRKLWALYLRDFFFKYTRQLESQ
KLELPIPLGLPTEKIVLLGASQRYVSGS"
BASE COUNT 461 a 268 c 347 g 407 t
ORIGIN
Query Match 11.5%; Score 193; DB 8; Length 1483;
Best Local Similarity 99.6%; Pred. No. 3.7e-95;
Matches 243; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
1 GGTAAAGCGTTTACTTATGTTTATATGCAACGGAAGATATGCGATTTGGAATGC 60
1226 GGTAAAGCGTTTACTTATGTTTATATGCAACGGAAGATATGCGATTTGGAATGC 1285
61 TTTTTCAGATCAATCAAGCTCCTACAGATTTCTTAGGGAATGTTTCAGGCTTTTGTTA 120
```

```
Db 1286 TTTTTCAGATCAATCAAGCTCCTACAGATTTCTTAGGGAATGTTTCAGGCTTTTGTTA 1345
2Y 121 GAAATTCGTTTATTGCAACAGGTAGAGACATAAACCATAGACAGATGATCTCGAAGAGA 180
Db 1346 GAAATTCGTTTATTGCAACAGGTAGAGACATAAACCATAGACAGATGATCTCGAAGAGA 1405
2Y 181 TAAGCTTCTCTATGCTTAAAGAAATGGACCGATACGAATAAAACAAGCATCATTAAGAT 240
Db 1406 TAAGCTTCTCTATGCTTAAAGAAATGGACCGATACGAATAAAACAAGCATCATTAAGAT 1465
2Y 241 TAAA 244
Db 1466 TAAA 1469

RESULT 4
LOCUS AY056214 1660 bp mRNA linear PLN 18-SEP-2002
DEFINITION Arabidopsis thaliana putative PCF2 DNA binding protein (AE2945680)
mRNA, complete cds.
ACCESSION AY056214
VERSION AY056214.1 GI:15810350
KEYWORDS FLI CDNA
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 1660)
AUTHORS Yamada, K., Liu, S.X., Sakano, H., Pham, P.K., Banh, J., Chung, M.K.,
Goldsmith, A.D., Lee, J.M., Quach, H.L., Toriumi, M., Yu, G., Bowser, L.,
Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T.,
Kamaya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B.,
Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T.,
Satou, M., Seki, M., Shinn, P., Southwick, A., Shinozaki, K.,
Davis, R.W., Ecker, J.R. and Theologis, A.
TITLE Arabidopsis Full Length cDNA Clones
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1660)
AUTHORS Yamada, K., Banh, J., Banno, F., Dale, J.M., Goldsmith, A.D., Lee, J.M.,
Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Yamamura, Y.,
Yu, G., Yu, S., Bowser, L., Carninci, P., Chen, H., Cheuk, R.,
Hayashizaki, Y., Ishida, J., Jones, T., Kamaya, A., Karlin-Neumann, G.,
Kawai, J., Kim, C., Koesema, E., Lam, B., Lin, J., Meyers, M.C.,
Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T.,
Satou, M., Seki, M., Shinn, P., Southwick, A., Tracy, S.E.,
Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.
TITLE Direct Submission
JOURNAL Submitted (12-SEP-2001) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
COMMENT RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAPL cDNAs (RAPL CDNA: 'RIKEN
Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J.,
Satou, M., Kamaya, A., Sakurai, T., Carninci, P., Kawai, J.,
Hayashizaki, Y. and Shinozaki, K.
The Salk, Stanford, PGEC (SSP) Consortium members carried out the
sequencing and annotation of the RAPL cDNAs: Yamada, K., Banh, J.,
Banno, F., Dale, J.M., Goldsmith, A.D., Lee, J.M., Onodera, C.S.,
Quach, H.L., Tang, C.C., Toriumi, M., Yamamura, Y., Yu, G., Yu, S.,
Bowser, L., Chen, H., Cheuk, R., Jones, T., Karlin-Neumann, G., Kim, C.,
Koesema, E., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Nguyen, M.,
Palm, C.J., Shinn, P., Southwick, A., Tracy, S.E., Davis, R.W.,
Ecker, J.R. and Theologis, A.
Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to
this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC)
contributed equally to this work as PIs.
Annotation is based on the January 2002 version of the Arabidopsis
Genome submitted to GenBank.
LOCATION/Qualifiers
1. .1660
FEATURES
source
```

```

/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="2"
/clone="RAFLO7-08-P04 (R10678)"
/note="This clone is in a modified pBluescript vector
(FUC-1) as a BamHI/XhoI insert.
ecotype: Columbia"
1..1660
/gene="At2g45680"
1..102
/gene="At2g45680"
103..1173
/gene="At2g45680"
/codon_start=1
/evidence="experimental"
/product="putative PCF2 DNA binding protein"
/protein_id="AA07063.1"
/db_xref="GI:15810351"
/translators="MATIOKLEVAGKOOTLRVDLTINGVNVTSPPFQNPVTS
LEKAEPMSEFSLAPSSGTPPLKRASTKDRHKVGRGRIRMPACARIFQL
TRELKSGDETRWLENAPFAIATATGTVFPAIMASVNGTKIPTINADSDMGE
NLMKRRKRNSEYIDISDAVSSGLAPIATITTIQPPALASSTVAQQLPQGM
PMWIPSNAMIPVGFAPFLPQAGSNOPQLAFPAASPSVVAQAQASTWARP
PIQVVPSSGQFVSDVSGSNLGRATSVNAPSSSGVTGSSSIATTTTLRDPSL
1174..1660
EIVEKELHOFMSTTTARSSNH"
1174..1660
/gene="At2g45680"
464 a 400 c 385 g 411 t
3'UTR
BASE COUNT
ORIGIN
Query Match 6.1%; Score 102; DB 8; Length 1660;
Best Local Similarity 100.0%; Pred. No. 1.4e-44;
Matches 102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
2Y 1576 CTTGGAGCTGTACCGCGTTAAACGATCTTCCCATCTGATCCGCTTTAACACTCT 1635
Db 1 CTTGGAGCTGTACCGCGTTAAACGATCTTCCCATCTGATCCGCTTTAACACTCT 60
2Y 1636 CTTGGAGCTGTACCGCGTTTCTCTCAGTATATTTTA 1677
Db 61 CTTGGAGCTGTACCGCGTTTCTCTCAGTATATTTTA 102
RESULT 5
AC106200/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-20C16, *** SEQUENCING IN PROGRESS
***, 3 unordered pieces.
AC106200
AC106200.4 GI:30579164
HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
Rattus norvegicus
SOURCE
Eukaryote; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
1 (bases 1 to 250957)
Muzny, D., Marie, E., Metzker, M., Lee, S., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blenkinsburg, K., Blyth, P., Brown, M.,
Bryant, N., Buay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Chen, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deranc, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,

```

```

Georgiadis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunnarsson, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, J.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hoque, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowals, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenz, H., Loulsged, H., Lozano, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapus, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Moran, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwakeleneh, O., Okwunonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plummer, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,
Puzos, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smas, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soza, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villaseca, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczyk, R., Woodden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, P., Zhang, J., Zhou, J., Zhou, X., Zhou, S., Yen, J., Yoon, L., Yoon, V.,
Niederhauser, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.
Direct Submission
Unpublished
2 (bases 1 to 250957)
Worley, K. C.
Direct Submission
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 250957)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 13, 2003 this sequence version replaced gi:23118059.
The sequence in this assembly is a combination of BAC-based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GJUV
Center clone name: CH230-20C16
----- Summary Statistics
Assembly program: Atlas 3.0
Consensus quality: 233323 bases at least Q40
Consensus quality: 235744 bases at least Q30

```

```

Consensus quality: 238092 bases at least Q20
Estimated insert size: 250301; sum-of-ctnigs estimation
Quality coverage: 6x in Q20 bases; sum-of-ctnigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 ctnigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the ctnigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 247058: contig of 247058 bp in length
* 247059 247158: gap of unknown length
* 247159 248228: contig of 1070 bp in length
* 248229 248328: gap of unknown length
* 248329 250957: contig of 2629 bp in length.
*
* FEATURES
*   source
*     1..250957
*       /organism="Rattus norvegicus"
*       /mol_type="genomic DNA"
*       /db_xref="taxon:10116"
*       /clone="CH230-20C16"
*
*   misc_feature
*     1..1145
*       /note="wgs_end_extension"
*       /clone_end="Sp6"
*       3949..4624
*       /note="clone_boundary"
*       /clone_end="Sp6"
*       site:EcoRI
*       end_sequence:BH318584"
*       complement(243703..244577)
*       /note="clone_boundary"
*       clone_end:T7
*       site:EcoRI
*       end_sequence:BH318581"
*       245116..247058
*       /note="wgs_end_extension"
*       clone_end:T7"
*
*   misc_feature
*     66339 49081 c 49937 g 74548 t 11052 others
*
* QUERY MATCH
*   Query Match 1.4%; Score 24; DB 2; Length 250957;
*   Best Local Similarity 100.0%; Pred.No.0.3;
*   Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
*
* Y 1097 CATAAATTAAGTAAATCTTTT 1120
*   |||||
* b 59625 CATAAATTAAGTAAATCTTTT 59602
*   |||||
*
* RESULT 6
* OCUS AC119331
* REFINITION Rattus norvegicus clone CH230-291E4, *** SEQUENCING IN PROGRESS
* ** 5 unordered pieces.
* CCESSION AC119331
* ERSION AC119331.5 GI:25099255
* EYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
* SOURCE Rattus norvegicus (Norway rat)
* ORGANISM Rattus norvegicus
*           Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
*           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
*           Rattus.
*
* REFERENCE 1 (bases 1 to 282156)
* AUTHORS Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,
* Allen, C., Allen, H., Allbrooks, S., Amin, A., Anguiano, D.,
* Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
* Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, P.,
* Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gburgeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, M., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hernandez, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, U., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hognes, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensheva, L., Loulbege, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Kwackelmech, O., Okwuonu, G., Olampunsaogon, A., Pal, S., Perez, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, O., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Speed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmami, K., Valas, R., Vera, V., Villanueva, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wisczyk, K., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs R.A.

Direct Submission
Unpublished
2 (bases 1 to 282156)
Worley, K.C.
Direct Submission
Submitted (26-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 282156)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 19, 2002 this sequence version replaced gi:23911491.
The sequence in this assembly is a combination of BAC-based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Center: Baylor College of Medicine

Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GUVU
 Center clone name: CH230-291E4
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 253046 bases at least Q40
 Consensus quality: 258512 bases at least Q30
 Consensus quality: 259010 bases at least Q20
 Estimated insert size: 261460; sum-of-contigs estimation
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
 * NOTE: This sequence may represent more than one clone.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 58359: contig of 58359 bp in length
 * 58360 58459: gap of unknown length
 * 58460 277096: contig of 218637 bp in length
 * 277097 277196: gap of unknown length
 * 277197 278294: contig of 1098 bp in length
 * 278295 278394: gap of unknown length
 * 278395 279912: contig of 1518 bp in length
 * 279913 280012: gap of unknown length
 * 280013 282156: contig of 2144 bp in length.
 *
 * Location/Qualifiers
 *
 * 1. 282156
 * /organism="Rattus norvegicus"
 * /mol_type="genomic DNA"
 * /db_xref="taxon:10116"
 * /clone="CH230-291E4"
 * 5162..6036
 * /note="clone boundary
 * clone_end:T7
 * site:
 * end sequence:BZ166106"
 * 58460..59573
 * /note="wgs contig"
 * complement(209107..209982)
 * /note="clone boundary
 * clone_end:Sp6
 * site:
 * end sequence:BZ166107"
 * 275379..277096
 * /note="wgs end extension
 * clone_end:Sp6"
 *
 * BASE COUNT 86729 a 49533 c 48244 g 75698 t 21962 others
 *
 * ORIGIN
 *
 * Query Match 1.4%; Score 24; DB 2; Length 282156;
 * Best Local Similarity 100.0%; Pred.No. 0.3;
 * Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 *
 * QY 1097 CATAAATTAAAGTAATCTTTT 1120
 *
 * Db 11519 CATAAATTAAAGTAATCTTTT 11542
 *
 * RESULT 7
 * LOCUS AX114617/c 101 bp DNA linear PAT 11-MAY-2001
 * DEFINITION Sequence 29 from Patent WO0128540.
 * ACCESSION AX114617
 * VERSION AX114617.1 GI:14031560
 *
 * SOURCE
 * Homo sapiens (human)
 * Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 * Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 *
 * REFERENCE
 * 1 Purvis, I.J. and McCarthy, L.C.
 * Therapy of cephalic pain
 * Patent: WO 0128540-A 29-26-APR-2001;
 * GLAXO GROUP LIMITED (GB)
 * LOCATION/Qualifiers
 * 1..101
 * /organism="Homo sapiens"
 * /mol_type="genomic DNA"
 * /db_xref="taxon:9606"
 * 26 a 15 c 14 g 46 t
 *
 * BASE COUNT 26 a 15 c 14 g 46 t
 *
 * ORIGIN
 *
 * Query Match 1.4%; Score 23; DB 6; Length 101;
 * Best Local Similarity 100.0%; Pred.No. 1.3;
 * Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 *
 * QY 223 ACAAGCATCATTAAGATTAAAT 245
 *
 * Db 75 ACAAGCATCATTAAGATTAAAT 53
 *
 * RESULT 8
 * LOCUS AX114618/c 101 bp DNA linear PAT 11-MAY-2001
 * DEFINITION Sequence 30 from Patent WO0128540.
 * ACCESSION AX114618
 * VERSION AX114618.1 GI:14031561
 *
 * SOURCE
 * Homo sapiens (human)
 * Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 * Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 *
 * REFERENCE
 * 1 Purvis, I.J. and McCarthy, L.C.
 * Therapy of cephalic pain
 * Patent: WO 0128540-A 30-26-APR-2001;
 * GLAXO GROUP LIMITED (GB)
 * LOCATION/Qualifiers
 * 1..101
 * /organism="Homo sapiens"
 * /mol_type="genomic DNA"
 * /db_xref="taxon:9606"
 * 27 a 14 c 14 g 46 t
 *
 * BASE COUNT 27 a 14 c 14 g 46 t
 *
 * ORIGIN
 *
 * Query Match 1.4%; Score 23; DB 6; Length 101;
 * Best Local Similarity 100.0%; Pred.No. 1.3;
 * Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 *
 * QY 223 ACAAGCATCATTAAGATTAAAT 245
 *
 * Db 75 ACAAGCATCATTAAGATTAAAT 53
 *
 * RESULT 9
 * LOCUS AX114743/c 101 bp DNA linear PAT 11-MAY-2001
 * DEFINITION Sequence 29 from Patent WO0128539.
 * ACCESSION AX114743
 * VERSION AX114743.1 GI:14031685
 *
 * SOURCE
 * Homo sapiens (human)
 * Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 * Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 *
 * REFERENCE
 * 1 Purvis, I.J. and McCarthy, L.C.

KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1
 REFERENCE Purvis, I.J. and McCarthy, L.C.
 AUTHORS Therapy of cephalic pain
 TITLE Therapy of cephalic pain
 JOURNAL Patent: WO 0128540-A 29-26-APR-2001;
 GLAXO GROUP LIMITED (GB)
 FEATURES Location/Qualifiers
 1..101
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 26 a 15 c 14 g 46 t
 source

BASE COUNT 26 a 15 c 14 g 46 t
 ORIGIN
 Query Match 1.4%; Score 23; DB 6; Length 101;
 Best Local Similarity 100.0%; Pred.No. 1.3;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ACAAGCATCATTAAGATTAAAT 245
 Db 75 ACAAGCATCATTAAGATTAAAT 53

RESULT 8
 LOCUS AX114618/c 101 bp DNA linear PAT 11-MAY-2001
 DEFINITION Sequence 30 from Patent WO0128540.
 ACCESSION AX114618
 VERSION AX114618.1 GI:14031561
 KEYWORDS
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1
 REFERENCE Purvis, I.J. and McCarthy, L.C.
 AUTHORS Therapy of cephalic pain
 TITLE Therapy of cephalic pain
 JOURNAL Patent: WO 0128540-A 30-26-APR-2001;
 GLAXO GROUP LIMITED (GB)
 FEATURES Location/Qualifiers
 1..101
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 27 a 14 c 14 g 46 t
 source

BASE COUNT 27 a 14 c 14 g 46 t
 ORIGIN
 Query Match 1.4%; Score 23; DB 6; Length 101;
 Best Local Similarity 100.0%; Pred.No. 1.3;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ACAAGCATCATTAAGATTAAAT 245
 Db 75 ACAAGCATCATTAAGATTAAAT 53

RESULT 9
 LOCUS AX114743/c 101 bp DNA linear PAT 11-MAY-2001
 DEFINITION Sequence 29 from Patent WO0128539.
 ACCESSION AX114743
 VERSION AX114743.1 GI:14031685
 KEYWORDS
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1
 REFERENCE Purvis, I.J. and McCarthy, L.C.

TITLE Agent for treating cephalic pain
JOURNAL Patent: WO 0128539-A 29 26-APR-2001;
GLAXO GROUP LIMITED (GB)

FEATURES

source

Location/Qualifiers
1. .101
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
26 a 15 c 14 g 46 t

BASE COUNT

ORIGIN

Query Match 1.4%; Score 23; DB 6; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2y 223 ACAAGCATCATTAAAGATTAAAT 245

Db 75 ACAAGCATCATTAAAGATTAAAT 53

RESULT 10

AX114744/c AX114744 101 bp DNA linear PAT 11-MAY-2001
LOCUS Sequence 30 from Patent WO0128539.
DEFINITION AX114744
ACCESSION AX114744
VERSION AX114744.1 GI:14031686
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1. .101
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
27 a 14 c 14 g 46 t

BASE COUNT

ORIGIN

Query Match 1.4%; Score 23; DB 6; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2y 223 ACAAGCATCATTAAAGATTAAAT 245

Db 75 ACAAGCATCATTAAAGATTAAAT 53

RESULT 11

AX118672/c AX118672 101 bp DNA linear PAT 11-MAY-2001
LOCUS Sequence 28 from Patent WO0129255.
DEFINITION AX118672
ACCESSION AX118672
VERSION AX118672.1 GI:14035623
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1. .101
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
26 a 15 c 14 g 46 t

BASE COUNT

ORIGIN

Query Match 1.4%; Score 23; DB 6; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2y 223 ACAAGCATCATTAAAGATTAAAT 245

Db 75 ACAAGCATCATTAAAGATTAAAT 53

RESULT 12

AX118673/c AX118673 101 bp DNA linear PAT 11-MAY-2001
LOCUS Sequence 29 from Patent WO0129255.
DEFINITION AX118673
ACCESSION AX118673
VERSION AX118673.1 GI:14035624
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1. .101
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
26 a 15 c 14 g 46 t

BASE COUNT

ORIGIN

Query Match 1.4%; Score 23; DB 6; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

BASE COUNT 26 a 15 c 14 g 46 t
ORIGIN

Query Match 1.4%; Score 23; DB 6; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 223 ACAAGCATCATTAAAGATTAAAT 245

Db 75 ACAAGCATCATTAAAGATTAAAT 53

RESULT 12

AX118673/c AX118673 101 bp DNA linear PAT 11-MAY-2001
LOCUS Sequence 29 from Patent WO0129255.
DEFINITION AX118673
ACCESSION AX118673
VERSION AX118673.1 GI:14035624
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1. .101
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
27 a 14 c 14 g 46 t

BASE COUNT

ORIGIN

Query Match 1.4%; Score 23; DB 6; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 223 ACAAGCATCATTAAAGATTAAAT 245

Db 75 ACAAGCATCATTAAAGATTAAAT 53

RESULT 13

AX128089/c AX128089 101 bp DNA linear PAT 15-MAY-2001
LOCUS Sequence 28 from Patent WO0129256.
DEFINITION AX128089
ACCESSION AX128089
VERSION AX128089.1 GI:14134629
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1. .101
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
26 a 15 c 14 g 46 t

BASE COUNT

ORIGIN

Query Match 1.4%; Score 23; DB 6; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 223 ACAAGCATCATTAAAGATTAAAT 245

Db 75 ACAAGCATCATTAAAGATTAAAT 53

RESULT 13

AX128089/c AX128089 101 bp DNA linear PAT 15-MAY-2001
LOCUS Sequence 28 from Patent WO0129256.
DEFINITION AX128089
ACCESSION AX128089
VERSION AX128089.1 GI:14134629
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1. .101
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
26 a 15 c 14 g 46 t

BASE COUNT

ORIGIN

Query Match 1.4%; Score 23; DB 6; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0;

Search completed: February 2, 2004, 07:55:32
Job time : 6370.7 secs

```
Qy 223 ACAAGCATCATTAAAGATTAAAT 245
Db 75 ACAAGCATCATTAAAGATTAAAT 53

RESULT 14
AX128090/c
LOCUS AX128090 101 bp DNA linear PAT 15-MAY-2001
DEFINITION Sequence 29 from Patent WO0129256.
ACCESSION AX128090
VERSION AX128090.1 GI:14134630
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Purvis, I.J. and McCarthy, L.C.
TITLE Cephalic pain susceptibility marker
JOURNAL Patent: WO 0129256-A 29 26-APR-2001;
GLAXO GROUP LIMITED (GB)
FEATURES
    source
        1..101
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
BASE COUNT 27 a 14 c 14 g 46 t
ORIGIN
Query Match 1.4%; Score 23; DB 6; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 23, Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 223 ACAAGCATCATTAAAGATTAAAT 245
Db 75 ACAAGCATCATTAAAGATTAAAT 53

RESULT 15
AX644690/c
LOCUS AX644690 101 bp DNA linear PAT 27-FEB-2003
DEFINITION Sequence 28 from Patent WO0233121.
ACCESSION AX644690
VERSION AX644690.1 GI:28610698
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Hosford, D. and Purvis, I.J.
TITLE Test for the diagnosis of diabetes and compounds for the treatment
JOURNAL Patent: WO 0233121-A 28 25-APR-2002;
GLAXO GROUP LIMITED (GB)
FEATURES
    source
        1..101
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
BASE COUNT 26 a 15 c 14 g 46 t
ORIGIN
Query Match 1.4%; Score 23; DB 6; Length 101;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 23, Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 223 ACAAGCATCATTAAAGATTAAAT 245
Db 75 ACAAGCATCATTAAAGATTAAAT 53
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M nucleic - nucleic search, using sw model

un on: February 1, 2004, 21:59:06 ; Search time 491.261 Seconds
(without alignments)
9214.976 Million cell updates/sec

itle: US-09-938-842A-3729

erfect score: 1677

equene: 1 ggtaagcgttttactatg.....tttctctcagctatatttta 1677

coring table: OLIGO NUC

Gapop 60.0 , Gapext 60.0

earched: 2552756 seqs, 1349719017 residues

ord size : 0

otal number of hits satisfying chosen parameters: 5105512

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Listing first 45 summaries

atabase : N Geneseq_19Jun03.*
1: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| result No. | Score | Query Match | Length | ID | Description |
|---------------|-------|----------------|--------|----|-------------|
| 1 | 1577 | 100.0 | 1677 | 24 | ABZ15924 |
| 2 | 193 | 11.5 | 1483 | 21 | AAH8782 |
| c 3 | 23 | 1.4 | 101 | 22 | AAH50475 |
| c 4 | 23 | 1.4 | 101 | 22 | AAH31172 |
| c 5 | 23 | 1.4 | 101 | 22 | AAH31269 |
| c 6 | 23 | 1.4 | 101 | 22 | AAH04594 |
| c 7 | 23 | 1.4 | 101 | 24 | ABQ72727 |
| c 8 | 23 | 1.4 | 7240 | 22 | AAH50570 |

| | | | | | | |
|------|----|-----|---------|----|----------|--------------------|
| c 9 | 23 | 1.4 | 7240 | 22 | AAH31170 | Human insulin rece |
| c 10 | 23 | 1.4 | 7240 | 22 | AAH31267 | Human insulin rece |
| c 11 | 23 | 1.4 | 7240 | 22 | AAH04467 | Human insulin rece |
| c 12 | 23 | 1.4 | 7240 | 24 | ABQ72725 | Human insulin rece |
| c 13 | 22 | 1.3 | 17646 | 25 | ABZ74645 | Secreted protein g |
| c 14 | 22 | 1.3 | 17646 | 25 | ABT17023 | Human secreted pro |
| c 15 | 22 | 1.3 | 17646 | 25 | ABZ68163 | Human secreted pro |
| c 16 | 22 | 1.3 | 1082138 | 21 | AAF22305 | Arabidopsis thalia |
| c 17 | 21 | 1.3 | 330 | 22 | AAH81545 | Human differential |
| c 18 | 21 | 1.3 | 394 | 22 | AAH37649 | Novel human diagno |
| c 19 | 21 | 1.3 | 407 | 22 | AAH38575 | Novel human diagno |
| c 20 | 21 | 1.3 | 1332 | 22 | AAH17158 | Human cDNA sequenc |
| c 21 | 21 | 1.3 | 2000 | 22 | AAH81797 | Human differential |
| c 22 | 21 | 1.3 | 2305 | 22 | AAH15783 | Human cDNA sequenc |
| c 23 | 21 | 1.3 | 2531 | 22 | AAH13793 | Human cDNA encodin |
| c 24 | 21 | 1.3 | 2625 | 23 | AAH82940 | DNA encoding novel |
| c 25 | 21 | 1.3 | 2704 | 21 | AAC76400 | Human ORFX ORF1955 |
| c 26 | 20 | 1.2 | 768 | 23 | AAH56350 | Salmonella typhi D |
| c 27 | 19 | 1.1 | 19 | 21 | AAH8786 | Arabidopsis thalia |
| c 28 | 19 | 1.1 | 22 | 21 | AAH8786 | Arabidopsis thalia |
| c 29 | 19 | 1.1 | 377 | 25 | ABX55749 | Bovine EST associa |
| c 30 | 19 | 1.1 | 400 | 18 | AAV78209 | Staphylococcus aur |
| c 31 | 19 | 1.1 | 406 | 21 | AAH30796 | Human colon cancer |
| c 32 | 19 | 1.1 | 408 | 22 | AAH28817 | Drosophila melanog |
| c 33 | 19 | 1.1 | 576 | 22 | AAH10030 | Human cDNA clone (|
| c 34 | 19 | 1.1 | 700 | 22 | AAH2907 | Human inflammatory |
| c 35 | 19 | 1.1 | 903 | 23 | AAH50398 | Staphylococcus aur |
| c 36 | 19 | 1.1 | 909 | 23 | AAH51756 | Staphylococcus aur |
| c 37 | 19 | 1.1 | 921 | 23 | AAH54565 | Haematopoietic cel |
| c 38 | 19 | 1.1 | 1630 | 25 | ABZ09980 | Arabidopsis thalia |
| c 39 | 19 | 1.1 | 2000 | 24 | ABZ16028 | Arabidopsis thalia |
| c 40 | 19 | 1.1 | 2352 | 24 | ABZ13787 | Arabidopsis thalia |
| c 41 | 19 | 1.1 | 2397 | 24 | ABZ14380 | Arabidopsis thalia |
| c 42 | 19 | 1.1 | 2500 | 23 | ABL16703 | Drosophila melanog |
| c 43 | 19 | 1.1 | 2925 | 23 | ABL08470 | Drosophila melanog |
| c 44 | 19 | 1.1 | 5011 | 23 | ABL15030 | Drosophila melanog |
| c 45 | 19 | 1.1 | 5901 | 24 | ABL70273 | Chemically treated |

ALIGNMENTS

RESULT 1
ABZ15924
ID ABZ15924 standard; DNA; 1677 BP.
XX
AC ABZ15924;
XX
DT 21-JAN-2003 (first entry)
XX
DE Arabidopsis thaliana stress regulated gene SEQ ID NO:3729.
XX
KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX
OS Arabidopsis thaliana.
XX
PN WC200216655-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26685.
XX
PR 24-AUG-2000; 2000US-227866P.
PR 26-JAN-2001; 2001US-264647P.
PR 22-JUN-2001; 2001US-300111P.
XX (SCRI) SCRIPPS RES INST.
XX (SYGN) SYNGENTA PARTICIPATIONS AG.
XX Harper JF, Krepis J, Wang X, Zhu T;
XX WPI; 2002-304127/34.

Identifying a stress condition to which a plant cell has been exposed and producing plants with increased tolerance to these abiotic stresses

Claim 144; SEQ ID NO 3729; 577bp + Sequence Listing; English.

The invention relates to identifying a stress condition to which a plant cell has been exposed, comprising:

(a) contacting nucleic acid representative of expressed polynucleotides in the plant cell with an array or probes representative of the plant cell genome; and

(b) detecting a profile of expressed polynucleotides in the plant cell characteristic of a stress response. The method is useful in the production of transgenic plants, cells and seeds and in producing plants with increased tolerance to abiotic stress. The present sequence is that of an Arabidopsis thaliana stress regulated gene (AB212196-AB21574) used in methods of the invention.

Note: The sequence data for this patent is not represented in the printed specification but is based on sequence information supplied to Derwent by the European Patent Office.

Sequence 1677 BP; 508 A; 292 C; 335 G; 542 T; 0 other;

Query Match 100.0%; Score 1677; DB 24; Length 1677;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1677; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GGTAAAGCGTTTACCTATGGTTTATATGCAAGGAGAAATATGCGATGTTGGATGC 60
 1 GGTAAAGCGTTTACCTATGGTTTATATGCAAGGAGAAATATGCGATGTTGGATGC 60
 61 TTTTTCAGATCATCAAGGCTCTACAGATTTCTTAGGGAATGTTTTCAGGCTTTTGTTA 120
 61 TTTTTCAGATCATCAAGGCTCTACAGATTTCTTAGGGAATGTTTTCAGGCTTTTGTTA 120
 121 GAAATGTTGTTTATGCAACAGGTAGAGAACATACATAGACAGATGATCTGAAGAGA 180
 121 GAAATGTTGTTTATGCAACAGGTAGAGAACATACATAGACAGATGATCTGAAGAGA 180
 181 TAAGCTTCTCTATGCTFAAAGAAATGACCGATACGAATAAACAAGCATCATTAAGAT 240
 181 TAAGCTTCTCTATGCTFAAAGAAATGACCGATACGAATAAACAAGCATCATTAAGAT 240
 241 TAAATGTTTGTGAAGAAATACATCTATTTATGTAATGTTGTTGTTAGTGAAGAT 300
 241 TAAATGTTTGTGAAGAAATACATCTATTTATGTAATGTTGTTGTTAGTGAAGAT 300
 301 AAAAACHATCGAATCCAAACCTCAATTTTACCAATCAGCCCAATTTATGATCTGGCG 360
 301 AAAAACHATCGAATCCAAACCTCAATTTTACCAATCAGCCCAATTTATGATCTGGCG 360
 361 TAATGAATGCTATGCTGATGTTAGGCAAAAGTTGGTGGCTGCGAAATTAACAACATTATC 420
 361 TAATGAATGCTATGCTGATGTTAGGCAAAAGTTGGTGGCTGCGAAATTAACAACATTATC 420
 421 CCTCTGTTGGACCCGAATCTGTAATCGGAAAGGTGGAACCCCACTGGTTTAACTTTT 480
 421 CCTCTGTTGGACCCGAATCTGTAATCGGAAAGGTGGAACCCCACTGGTTTAACTTTT 480
 481 AAGCCTAAAGAGTTACTACCGGTTTACCGGTTTATATTTTGGTGTGTTAATCTAATCC 540
 481 AAGCCTAAAGAGTTACTACCGGTTTACCGGTTTATATTTTGGTGTGTTAATCTAATCC 540
 541 GGATCCGTTTGTGTTTAAATCTCAAGGCCACGTTATCGCCCAATTTTGAATTTTGGTG 600
 541 GGATCCGTTTGTGTTTAAATCTCAAGGCCACGTTATCGCCCAATTTTGAATTTTGGTG 600
 601 GGTAGGAATGTTGGGTGGAATAGTTTGGGCTTAGCCCTCAACAATGTGGAACATGAG 660
 601 GGTAGGAATGTTGGGTGGAATAGTTTGGGCTTAGCCCTCAACAATGTGGAACATGAG 660
 661 AGAGTAGGGTCCAGCTCAGGCCACATTCACATTTTGGTTTCTAGCCCTTCTTTTCTGG 720
 661 AGAGTAGGGTCCAGCTCAGGCCACATTCACATTTTGGTTTCTAGCCCTTCTTTTCTGG 720

DB 661 AGAGTAGGGTCCAGCTCAGGCCACATTCACATTTTGGTTTCTAGCCCTTCTTTTCTGG 720
 QY 721 TGCATTACCGTCCCTCTTTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 780
 DB 721 TGCATTACCGTCCCTCTTTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 780
 QY 781 CCCGAAACAGTACCAACGAAATCAAAATTAAGTTTGAATCGGTTACATCTAGTTACCGTGG 840
 DB 781 CCCGAAACAGTACCAACGAAATCAAAATTAAGTTTGAATCGGTTACATCTAGTTACCGTGG 840
 QY 841 AACTTACATCATTTTCGATTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 900
 DB 841 AACTTACATCATTTTCGATTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 900
 QY 901 CCGATTGTCAAGTACCAAGTACATAGTATGCGTATATGATGATGATGATGATGATGATGATGAT 960
 DB 901 CCGATTGTCAAGTACCAAGTACATAGTATGCGTATATGATGATGATGATGATGATGATGATGAT 960
 QY 961 AAAGGACGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
 DB 961 AAAGGACGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
 QY 1021 ATTAGCTTTTGGTGGCGAGCCTTGTGACCTACTTAAATGCGGTCCAAACCCCAAGTATG 1080
 DB 1021 ATTAGCTTTTGGTGGCGAGCCTTGTGACCTACTTAAATGCGGTCCAAACCCCAAGTATG 1080
 QY 1081 GGCCTACAGCTTTTCCATAAATTAAGTAAATCTTTTTCGCTAAACCAATAAAAAT 1140
 DB 1081 GGCCTACAGCTTTTCCATAAATTAAGTAAATCTTTTTCGCTAAACCAATAAAAAT 1140
 QY 1141 ATTGAAATCTTTCCAAACCATAGAAAGTTTAAATTTGATCAGCGATGGAATTTTGTAC 1200
 DB 1141 ATTGAAATCTTTCCAAACCATAGAAAGTTTAAATTTGATCAGCGATGGAATTTTGTAC 1200
 QY 1201 AAAGTACGTTTTCATTTGCGGAGTGTACTAGTAACTAGTAACTAGTAACTAGTAACTAGTAA 1260
 DB 1201 AAAGTACGTTTTCATTTGCGGAGTGTACTAGTAACTAGTAACTAGTAACTAGTAACTAGTAA 1260
 QY 1261 TTTCTGATTTTGGATTTTGAAGCTTTTCTTAGGTTTAAAAAACAAGTATATTTACTTAAACAAT 1320
 DB 1261 TTTCTGATTTTGGATTTTGAAGCTTTTCTTAGGTTTAAAAAACAAGTATATTTACTTAAACAAT 1320
 QY 1321 AAAAGAAAAACATTTTGTGAAAAAGAGAAATTAAGTTTACTTGGACCCCATTTGACAGATGG 1380
 DB 1321 AAAAGAAAAACATTTTGTGAAAAAGAGAAATTAAGTTTACTTGGACCCCATTTGACAGATGG 1380
 QY 1381 TCCCATTAATTAATCTGATCAAGATAGAGCAATGGAAGTGTGTTGTTGTTGTTGTTGTTGTTGTT 1440
 DB 1381 TCCCATTAATTAATCTGATCAAGATAGAGCAATGGAAGTGTGTTGTTGTTGTTGTTGTTGTTGTT 1440
 QY 1441 TCGGAATGTTTCTTTAAAGCTCATCGAAACATCAGACCGTTGATTTTCCCGCATCAA 1500
 DB 1441 TCGGAATGTTTCTTTAAAGCTCATCGAAACATCAGACCGTTGATTTTCCCGCATCAA 1500
 QY 1501 AAACGTTTGAATTAATCTCACTGTTTCTGCTGCTTATATATATATATATATATATATATATATAT 1560
 DB 1501 AAACGTTTGAATTAATTAATCTCACTGTTTCTGCTGCTTATATATATATATATATATATATATAT 1560
 QY 1561 ACATTTAGTAAATCTCTGCGATGTAACCGCGTTAAACCGATTTCTTTCCCATTTGATCC 1620
 DB 1561 ACATTTAGTAAATCTCTGCGATGTAACCGCGTTAAACCGATTTCTTTCCCATTTGATCC 1620
 QY 1621 GCTTTTAAACCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1677
 DB 1621 GCTTTTAAACCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1677

RESULT 2
 AA88782
 ID AA88782 standard; cDNA; 1483 BP.
 XX
 AC AA88782;
 XX

T 19-FEB-2001 (first entry)
 X Arabidopsis thaliana SSEE1 cDNA.
 E SSEE1; shrunken seed gene; storage reserve; storage protein;
 W oil body; transgenic plant; ss.
 W Arabidopsis thaliana.
 S
 X
 H Key Location/Qualifiers
 H 122..1225
 T CDS /*tag= a
 T
 X WO200061735-A1.
 N 19-OCT-2000.
 D
 D 07-APR-2000; 2000WO-US09192.
 F
 F 08-APR-1999; 99US-0128651.
 R (GEHO) GEN HOSPITAL CORP.
 A
 X Lin Y;
 I
 X WPI: 2000-679483/66.
 R P-PSDB; AAB19718.
 R
 T Novel shrunken seed gene useful for producing transgenic plants having
 T altered production of food storage reserve material, intracellular
 T transport of storage protein and formation of protein or oil bodies -
 S Claim 8; Page 57; 64pp; English.
 X
 X The present sequence is that of Arabidopsis thaliana SSEE1 (shrunken
 C seed) cDNA, which was isolated from a seedling cDNA library using
 C a partial genomic clone as probe. SSEE1 encodes a protein (see
 C AAB19718) that, when expressed in a cell of a plant, modifies or
 C alters the production of a food storage reserve material (e.g.
 C protein, lipid or carbohydrate storage reserve), facilitates the
 C intracellular transport of a storage protein, or facilitates the
 C formation of protein or oil bodies. The invention provides a
 C transgenic plant (or plant cell, plant tissue, plant organ or
 C plant component) which includes a recombinant SSEE1 transgene that
 C modifies the production of food storage reserves, thereby
 C increasing nutritional value. An antisense construct is useful for
 C modifying desiccation tolerance.
 X
 X Sequence 1483 BP; 461 A; 268 C; 347 G; 407 T; 0 other;
 Q
 Query Match 11.5%; Score 193; DB 21; Length 1483;
 Best Local Similarity 99.6%; Pred No. 5, 9e-85;
 Matches 243; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Y 1 GGTAAAGCGTTTACTTATGGTTTATATGCAACGGAATATTTGCCATTTGGAATGC 60
 b 1226 GGTAAAGCGTTTACTTATGGTTTATATGCAACGGAATATTTGCCATTTGGAATGC 1285
 Y 61 TTTTTCAGATCATCAAGGCTTCACAGATTTCTTAGGGAATGTTTCAGGCTTTTGTGA 120
 b 1286 TTTTTCAGATCATCAAGGCTTCACAGATTTCTTAGGGAATGTTTCAGGCTTTTGTGA 1345
 Y 121 GAAATGTGTTTATTGCAACAGGTAGAGAAACATACCATAGACAGATGTATCTCGAAGAGA 180
 b 1346 GAAATGTGTTTATTGCAACAGGTAGAGAAACATACCATAGACAGATGTATCTCGAAGAGA 1405
 Y 181 TTAGCTTCTATGCTTAAGAAATGGCCCATAGCAATTAACACGATCATTTAAAGAT 240
 b 1406 TTAGCTTCTATGCTTAAGAAATGGCCCATAGCAATTAACACGATCATTTAAAGAT 1465
 Y 241 TAAA 244
 b 1466 TAAA 1469

RESULT 3

AAH50475/c
 ID AAH50475 standard; DNA; 101 BP.
 XX
 AC AAH50475;
 XX
 XX 21-AUG-2001 (first entry)
 DT
 XX Insulin receptor migraine associated polymorphic site #2.
 DE
 XX Insulin receptor; polymorphic site; single nucleotide polymorphism;
 KW SNP; migraine; cephalic pain; insulin receptor signalling pathway;
 KW antimigraine; vulnery; cluster headache; chronic paroxysmal hemicrania;
 KW vascular disorder associated headache; withdrawal; tension headache; ds.
 XX
 OS Homo sapiens.
 XX
 XX Location/Qualifiers
 FH Key replace(51,A)
 FT allele /*tag= a
 FT /standard_name= "single nucleotide polymorphism"
 FT
 XX WO200128539-A2.
 PN 26-APR-2001.
 PD
 XX 19-OCT-2000; 2000WO-GB04031.
 PF
 XX 19-OCT-1999; 99GB-0024713.
 PR
 PR 19-OCT-1999; 99US-0160423.
 XX
 XX (GLAX) GLAXO GROUP LTD.
 PA
 XX Purvis IJ, McCarthy LC;
 PI WPI; 2001-300274/31.
 XX
 XX Use of agent that modulates directly or indirectly insulin receptor or
 PT insulin receptor signaling pathway in the manufacture of medicament for
 PT preventing or treating cephalic pain -
 XX Example; Page 34; 58pp; English.
 XX The present invention describes the use of an agent (I) that modulates
 CC directly or indirectly the insulin receptor or insulin receptor
 CC signalling pathway in the manufacture of a medicament for preventing or
 CC treating cephalic pain. Also described is an isolated polynucleotide (II)
 CC or protein (III) comprising a polymorphism that causes susceptibility to
 CC cephalic pain, or a naturally occurring polymorphism that is in linkage
 CC disequilibrium with the first polymorphism. (I) has antimigraine and
 CC vulnery activities. (I) is useful for treating cephalic pain which may
 CC be a cluster headache, chronic paroxysmal hemicrania, headache
 CC associated with vascular disorders, headache associated with substances
 CC or their withdrawal (for example drug withdrawal), tension headache and
 CC in particular migraine with or without aura. The present sequence
 CC represents an insulin receptor oligonucleotide containing a migraine
 CC associated polymorphic site, which is used in the exemplification of
 CC the present invention.
 CC
 SQ Sequence 101 BP; 26 A; 15 C; 14 G; 46 T; 0 other;
 Query Match 1.4%; Score 23; DB 22; Length 101;
 Best Local Similarity 100.0%; Pred. No. 0.69;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 223 ACAAGCATCATTTAAAGATTAAAT 245
 Db 75 ACAAGCATCATTTAAAGATTAAAT 53

RESULT 4

AAH31172/c
ID AAH31172 standard; DNA; 101 BP.

XX AC

AAH31172;

XX DT 25-JUL-2001 (first entry)

XX DE Human insulin receptor gene polymorphic site INSBa.

XX KW Human; insulin receptor; diagnosis; cephalic pain; susceptibility;
single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

KW migraine; ds.

XX OS Homo sapiens.

XX FH Key

FT variation

FT Location/Qualifiers

FT replace(51,A)

FT /*tag= a

FT /standard_name= "Single nucleotide polymorphism"

XX WO200129255-A2.

XX PD 26-APR-2001.

XX PF 19-OCT-2000; 2000WO-GB04050.

XX PR 19-OCT-1999; 99GB-0024717.

XX PA (GLAX) GLAXO GROUP LTD.

XX PI Purvis IJ, McCarthy LC;

XX DR WPI; 2001-316247/33.

XX PT Diagnosing susceptibility to cephalic pain such as migraine by typing

XX PT insulin receptor gene or protein in vivo, or in a biological sample and

XX PT determining individual's susceptibility to cephalic pain

XX PS Example; Page 21; 46pp; English.

XX CC The present sequence is provided in a specification relating to a

XX CC method for diagnosing susceptibility to cephalic pain in an individual.

XX CC The method comprises typing in vivo or in a sample from the

XX CC individual, the insulin receptor gene region or insulin receptor

XX CC protein and thus determining whether the individual is susceptible to

XX CC cephalic pain. Polymorphisms in the insulin receptor gene that

XX CC cause susceptibility to cephalic pain have been found. Susceptibility

XX CC may be diagnosed using a probe, primer or antibody which is capable of

XX CC detecting an insulin receptor gene region or insulin receptor protein

XX CC polymorphism. The method is useful for diagnosing susceptibility to

XX CC migraine, cluster headache, chronic paroxysmal hemianopia, headache

XX CC associated with vascular disorders, headache associated with substances

XX CC or their withdrawal, tension headache, and so on. It is useful for

XX CC assessing the efficacy of agents in relieving cephalic pain, and

XX CC can be used to assess the ability of agents to modulate insulin

XX CC receptor signalling activity. The method may also be used to assess

XX CC the predisposition and/or susceptibility of an individual to the

XX CC development of diseases mediated by the insulin receptor and in the

XX CC development of new drug therapies which selectively target one or more

XX CC allelic variants of the insulin receptor gene. The present sequence

XX CC corresponds to a polymorphic site and flanking sequences of a single

XX CC nucleotide polymorphism (SNP) in the human insulin receptor gene.

XX SQ Sequence 101 BP; 26 A; 15 C; 14 G; 46 T; 0 other;

Query Match 1.4%; Score 23; DB 22; Length 101;

Best Local Similarity 100.0%; Pred. No. 0.69;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ACAAGCATCATTAAGATTAAAT 245

|||||

DB 75 ACAAGCATCATTAAGATTAAAT 53

|||||

RESULT 6

AA04594/c

ID AA04594 standard; DNA; 101 BP.

XX AC

AA04594;

XX DT 04-JUL-2001 (first entry)

XX DE INSBa polymorphism of human insulin receptor gene.

XX KW

AAH31172/c

ID AAH31172 standard; DNA; 101 BP.

XX AC

AAH31172;

XX DT 25-JUL-2001 (first entry)

XX DE Human insulin receptor gene polymorphic site INSBa.

XX KW Human; insulin receptor; diagnosis; cephalic pain; susceptibility;
single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

XX KW single nucleotide polymorphism; SNP; migraine; SNP detection; ds.

W Human; insulin receptor; cephalic pain; therapy; headache;
W chronic paroxysmal hemiparesis; vascular disorder; tension headache;
W migraine; single nucleotide polymorphism; SNP; ds.
S Homo sapiens.

X Key Location/Qualifiers
H replace (S1, A)
I /*tag= a
T /standard_name= "single nucleotide polymorphism"

T WO200128540-A2.

N 26-APR-2001.

D 19-OCT-2000; 2000WO-GB04051.

F 19-OCT-1999; 99GB-0024712.

R 19-OCT-1999; 99US-0160418.

X (GLAX) GLAXO GROUP LTD.

X Purvis IJ, McCarthy LC;

I WPI; 2001-290815/30.

R Use of agent that modulates directly or indirectly insulin receptor or
R insulin receptor signaling pathway in manufacture of medicament for
R preventing or treating cephalic pain -
S Claim 7; Page 21; 46pp; English.

X The present invention relates to the use of an agent that modulates
X directly or indirectly the insulin receptor or insulin receptor
X signalling pathway in the manufacture of a medicament for preventing
X or treating cephalic pain. Cephalic pain disorders are generally
X multifunctional disorder, with a cluster of headache, chronic
X paroxysmal hemiparesis, headache associated with vascular disorders,
X headache associated with substances and their withdrawal (for example
X drug withdrawal), tension headache and in particular migraine with aura
X or migraine without aura. The treatment of cephalic pain and migraine
X involves manipulation of components of the glucose and lipid metabolism
X pathways, in particular by manipulation of the insulin receptor. Insulin
X receptor is an important component in the regulation of the glucose and
X lipid metabolism pathways. Single nucleotide polymorphisms (SNP) in the
X insulin receptor gene contribute to susceptibility to cephalic pain.
X The present sequence is INSBA polymorphism of human insulin receptor
X gene.

X Sequence 101 BP; 26 A; 15 C; 14 G; 46 T; 0 other;

X Query Match 1.4%; Score 23; DB 22; Length 101;

X Best Local Similarity 100.0%; Pred. No. 0.69;

X Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 223 ACAAGCATCATTAAGATTAAAT 245

b 75 ACAAGCATCATTAAGATTAAAT 53

RESULT 7

EC72727/c

D ABQ72727 standard; DNA; 101 BP.

X ABQ72727;

X 09-SEP-2002 (first entry)

E Human insulin receptor gene polymorphism INSBA.

X Human; insulin; receptor; diabetes; antidiabetic; ds;
W single nucleotide polymorphism; SNP.

OS Homo sapiens.
XX Key Location/Qualifiers
FH variation replace (S1, A)
FT /*tag= a
FT /standard_name= "single nucleotide polymorphism"

PN WO200233121-A2.

XX 25-APR-2002.

XX 19-OCT-2001; 2001WO-GB04660.

XX 19-OCT-2000; 2000GB-0025678.

XX (GLAX) GLAXO GROUP LTD.

XX Hosford D, Purvis IJ;

XX WPI; 2002-500014/53.

XX Diagnosing diabetes or susceptibility to diabetes in individual by
PT typing insulin receptor gene region or insulin receptor protein in
PT sample obtained from individual -
XX Claim 5; Page 31; 61pp; English.

XX he invention relates to a novel method for diagnosing diabetes or

CC susceptibility to diabetes in an individual. The method of the invention
CC has antidiabetic activity. The method is useful for diagnosing diabetes
CC or susceptibility to diabetes in an individual. Other methods of the
CC invention are useful for treating diabetes and for treating and
CC preventing diabetes. The present sequence contains a single nucleotide
CC polymorphism (SNP) present in the human insulin receptor gene.

XX Sequence 101 BP; 26 A; 15 C; 14 G; 46 T; 0 other;

XX Query Match 1.4%; Score 23; DB 24; Length 101;

XX Best Local Similarity 100.0%; Pred. No. 0.69;

XX Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 223 ACAAGCATCATTAAGATTAAAT 245

Db 75 ACAAGCATCATTAAGATTAAAT 53

RESULT 8

AAH50570/c

ID AAH50570 standard; DNA; 7240 BP.

XX AAH50570;

XX 21-AUG-2001 (first entry)

XX Insulin receptor gene exons 14-17 including introns SEQ ID NO:25.

XX Insulin receptor; polymorphic site; single nucleotide polymorphism;
KW SNP; migraine; cephalic pain; insulin receptor signalling pathway;
KW antihypertensive; vulnery; cluster headache; chronic paroxysmal hemiparesis;
KW vascular disorder associated headache; withdrawal; tension headache; ds.

OS Homo sapiens.

XX WO200128539-A2.

XX 26-APR-2001.

XX 19-OCT-2000; 2000WO-GB04031.

XX 19-OCT-1999; 99GB-0024713.

XX 19-OCT-1999; 99US-0160423.

XX (GLAX) GLAXO GROUP LTD.

CC in the insulin receptor gene have been found to cause susceptibility to
CC cephalic pain, particularly to migraine. The method comprises typing
CC a region of the insulin receptor gene or insulin receptor protein of
CC an individual. A susceptibility to cephalic pain may be diagnosed
CC using a probe, primer or antibody which is capable of detecting a
CC polymorphism in an insulin receptor gene region or the insulin
CC receptor protein.
XX
SQ Sequence 7240 BP; 1952 A; 1653 C; 1820 G; 1815 T; 0 other;
Query Match 1.4%; Score 23; DB 22; Length 7240;
Best Local Similarity 100.0%; Pred. No. 0.56;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 223 ACAAGCATCATTAAGATTAAAT 245
DB 5056 ACAAGCATCATTAAGATTAAAT 5034
RESULT 10
AAH31267/c
ID AAH31267 standard; DNA; 7240 BP.
XX
AC AAH31267;
DT 25-JUL-2001 (first entry)
XX
XX Human insulin receptor gene from exon 14 to 17.
DE
XX
XX Human; insulin receptor; antimigraine; cephalic pain; diagnosis;
KW susceptibility; single nucleotide polymorphism; SNP; SNP detection;
KW migraine; ds.
XX
XX Homo sapiens.
OS
XX
FN WO200129255-A2.
XX
PD 26-APR-2001.
XX
XX 19-OCT-2000; 2000WO-GB04050.
PF
XX
PR 19-OCT-1999; 99GB-0024717.
XX
XX (GLAX) GLAXO GROUP LTD.
XX
XX Purvis IJ, McCarthy LC;
PI
XX
XX WPI; 2001-316247/33.
DR
XX
PT Diagnosing susceptibility to cephalic pain such as migraine by typing
PT insulin receptor gene or protein in vivo, or in a biological sample and
PT determining individual's susceptibility to cephalic pain -
XX
XX Disclosure; Page 43-46; 46pp; English.
XX
XX The present sequence is provided in a specification relating to a
CC method for diagnosing susceptibility to cephalic pain in an individual.
CC The method comprises typing in vivo or in a sample from the
CC individual, the insulin receptor gene region or insulin receptor
CC protein and thus determining whether the individual is susceptible to
CC cephalic pain. Polymorphisms in the insulin receptor gene that
CC cause susceptibility to cephalic pain have been found. Susceptibility
CC may be diagnosed using a probe, primer or antibody which is capable of
CC detecting an insulin receptor gene region or insulin receptor protein
CC polymorphism. The method is useful for diagnosing susceptibility to
CC migraine, cluster headache, chronic paroxysmal hemicrania, headache
CC associated with vascular disorders, headache associated with substances
CC or their withdrawal, tension headache, and so on. It is useful for
CC assessing the efficacy of agents in relieving cephalic pain, and
CC can be used to assess the ability of agents to modulate insulin
CC receptor signalling activity. The method may also be used to assess
CC the predisposition and/or susceptibility of an individual to the
CC development of diseases mediated by the insulin receptor and in the

CC in the insulin receptor gene have been found to cause susceptibility to
CC cephalic pain, particularly to migraine. The method comprises typing
CC a region of the insulin receptor gene or insulin receptor protein of
CC an individual. A susceptibility to cephalic pain may be diagnosed
CC using a probe, primer or antibody which is capable of detecting a
CC polymorphism in an insulin receptor gene region or the insulin
CC receptor protein.
XX
SQ Sequence 7240 BP; 1952 A; 1653 C; 1820 G; 1815 T; 0 other;
Query Match 1.4%; Score 23; DB 22; Length 7240;
Best Local Similarity 100.0%; Pred. No. 0.56;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 223 ACAAGCATCATTAAGATTAAAT 245
DB 5056 ACAAGCATCATTAAGATTAAAT 5034
RESULT 10
AAH31267/c
ID AAH31267 standard; DNA; 7240 BP.
XX
AC AAH31267;
DT 25-JUL-2001 (first entry)
XX
XX Human insulin receptor gene from exon 14 to 17.
DE
XX
XX Human; insulin receptor; antimigraine; cephalic pain; diagnosis;
KW susceptibility; single nucleotide polymorphism; SNP; SNP detection;
KW migraine; ds.
XX
XX Homo sapiens.
OS
XX
FN WO200129255-A2.
XX
PD 26-APR-2001.
XX
XX 19-OCT-2000; 2000WO-GB04050.
PF
XX
PR 19-OCT-1999; 99GB-0024717.
XX
XX (GLAX) GLAXO GROUP LTD.
XX
XX Purvis IJ, McCarthy LC;
PI
XX
XX WPI; 2001-316247/33.
DR
XX
PT Diagnosing susceptibility to cephalic pain such as migraine by typing
PT insulin receptor gene or protein in vivo, or in a biological sample and
PT determining individual's susceptibility to cephalic pain -
XX
XX Disclosure; Page 43-46; 46pp; English.
XX
XX The present sequence is provided in a specification relating to a
CC method for diagnosing susceptibility to cephalic pain in an individual.
CC The method comprises typing in vivo or in a sample from the
CC individual, the insulin receptor gene region or insulin receptor
CC protein and thus determining whether the individual is susceptible to
CC cephalic pain. Polymorphisms in the insulin receptor gene that
CC cause susceptibility to cephalic pain have been found. Susceptibility
CC may be diagnosed using a probe, primer or antibody which is capable of
CC detecting an insulin receptor gene region or insulin receptor protein
CC polymorphism. The method is useful for diagnosing susceptibility to
CC migraine, cluster headache, chronic paroxysmal hemicrania, headache
CC associated with vascular disorders, headache associated with substances
CC or their withdrawal, tension headache, and so on. It is useful for
CC assessing the efficacy of agents in relieving cephalic pain, and
CC can be used to assess the ability of agents to modulate insulin
CC receptor signalling activity. The method may also be used to assess
CC the predisposition and/or susceptibility of an individual to the
CC development of diseases mediated by the insulin receptor and in the

development of new drug therapies which selectively target one or more allelic variants of the insulin receptor gene.

Sequence 7240 BP; 1952 A; 1653 C; 1820 G; 1815 T; 0 other;

Query Match 1.4%; Score 23; DB 22; Length 7240;

Best Local Similarity 100.0%; Pred. No. 0.56; Indels 0; Gaps 0;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

223 ACAAGCATCATTAAGATTAAAT 245

|||||

5056 ACAAGCATCATTAAGATTAAAT 5034

RESULT 11

ID04467/c

AAD04467 standard; DNA; 7240 BP.

AAD04467;

04-JUL-2001 (first entry)

Human insulin receptor DNA (from exons 14 to 17).

Human; insulin receptor; cephalic pain; therapy; headache;
chronic paroxysmal hemiparesis; vascular disorder; tension headache;
migraine; single nucleotide polymorphism; SNP; ds.

Homo sapiens.

WO200128540-A2.

26-APR-2001.

19-OCT-2000; 2000WO-GB04051.

19-OCT-1999; 99GB-0024712.

19-OCT-1999; 99US-0160418.

(GLAX) GLAXO GROUP LTD.

Purvis IJ, McCarthy LC;

WPI; 2001-290815/30.

Use of agent that modulates directly or indirectly insulin receptor or insulin receptor signaling pathway in manufacture of medicament for preventing or treating cephalic pain -

Disclosure; Page 42-45; 46pp; English.

The present invention relates to the use of an agent that modulates directly or indirectly the insulin receptor or insulin receptor signalling pathway in the manufacture of a medicament for preventing or treating cephalic pain. Cephalic pain disorders are generally multifunctional disorder, with a cluster of headache, chronic paroxysmal hemiparesis, headache associated with vascular disorders, headache associated with substances or their withdrawal (for example drug withdrawal), tension headache and in particular migraine with aura or migraine without aura. The treatment of cephalic pain and migraine involves manipulation of components of the glucose and lipid metabolism pathways, in particular by manipulation of the insulin receptor. Insulin receptor is an important component in the regulation of the glucose and lipid metabolism pathways. Single nucleotide polymorphisms (SNP) in the insulin receptor gene contribute to susceptibility to cephalic pain. The present sequence is complete sequence from exon 14 to 17 of human insulin receptor gene. SNP in this sequence contributes to susceptibility to cephalic pain.

Sequence 7240 BP; 1952 A; 1653 C; 1820 G; 1815 T; 0 other;

Query Match

Best Local Similarity 1.4%; Score 23; DB 22; Length 7240;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ACAAGCATCATTAAGATTAAAT 245

|||||

Db 5056 ACAAGCATCATTAAGATTAAAT 5034

RESULT 12

ABQ72725/c

ID ABQ72725 standard; DNA; 7240 BP.

XX AC ABQ72725;

XX 09-SEP-2002 (first entry)

XX Human insulin receptor gene.

XX Human; insulin; receptor; diabetes; antidiabetic; ds.

XX Homo sapiens.

XX WO200233121-A2.

XX 25-APR-2002.

XX 19-OCT-2001; 2001WO-GB04660.

XX 19-OCT-2000; 2000GB-0025678.

XX (GLAX) GLAXO GROUP LTD.

XX Hosford D, Purvis IJ;

XX WPI; 2002-500014/53.

XX Diagnosing diabetes or susceptibility to diabetes in individual by typing insulin receptor gene region or insulin receptor protein in sample obtained from individual -

XX Disclosure; Page 57-61; 61pp; English.

XX he invention relates to a novel method for diagnosing diabetes or susceptibility to diabetes in an individual. The method of the invention has antidiabetic activity. The method is useful for diagnosing diabetes or susceptibility to diabetes in an individual. Other methods of the invention are useful for treating diabetes and for treating and preventing diabetes. The present sequence represents the human insulin receptor gene.

XX Sequence 7240 BP; 1952 A; 1653 C; 1820 G; 1815 T; 0 other;

Query Match 1.4%; Score 23; DB 24; Length 7240;

Best Local Similarity 100.0%; Pred. No. 0.56;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 ACAAGCATCATTAAGATTAAAT 245

|||||

Db 5056 ACAAGCATCATTAAGATTAAAT 5034

RESULT 13

ABZ74645/c

ID ABZ74645 standard; DNA; 17646 BP.

XX AC ABZ74645;

XX 12-MAY-2003 (first entry)

XX Secreted protein gene 381 genomic fragment HE2CA60, SEQ ID NO:1792.

XX Human; secreted protein; cancer; tumour; hyperproliferative disorder;
XX autoimmune disorder; inflammation; angiogenic diseases; AIDS;
XX acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;

drug screening; chromosome identification; chromosome mapping; cytostatic; gene therapy; anti-inflammatory; immunomodulator; anti-HIV; antianemic; vulnery; gene; ds.

Homo sapiens.

WO200277013-A2.

03-OCT-2002.

26-MAR-2002; 2002WO-US09370.

27-MAR-2001; 2001US-278650P.

12-SEP-2001; 2001US-0950082.

12-SEP-2001; 2001US-0950083.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Ruben SM;

WPI; 2003-040578/03.

New human secreted proteins and nucleic acids, useful for detecting or treating cancer or other hyperproliferative disorders, autoimmune disorders, inflammatory disorders, HIV disease, hepatitis or anemia - Disclosure; Page 2382-2387; 2474pp; English.

ABZ73281-ABZ73697 represent cDNAs corresponding to 391 human secreted protein genes, and ABP00947-ABP01363 represent the proteins they encode. The invention also encompasses antibodies specific for the secreted proteins, the use of the secreted proteins in drug screening and recombinant vectors and host cells comprising a nucleic acid of the invention. The secreted proteins are thought to be involved in biological activities associated with cellular signalling, cellular differentiation, cell migration, prothrombin activation and neurotransmitter activity. The secreted proteins, nucleic acids encoding them, antibodies or antibody fragments specific for the secreted proteins, and modulators of protein activity are useful for diagnosing or treating cancers or other hyperproliferative disorders. Additionally, the secreted proteins and their nucleic acids may also be used in the treatment of autoimmune disorders, inflammatory disorders, diseases involving angiogenesis, AIDS (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote wound healing. Nucleic acids of the invention may be used for chromosome identification, chromosome mapping, in gene therapy, for identifying individuals from minute biological samples, as hybridisation probes, and as molecular weight markers. The present sequence represents a human secreted protein genomic fragment referred to in the disclosure of the invention.

Sequence 17646 BP; 5370 A; 3099 C; 3561 G; 5616 T; 0 other;

Query Match 1.3%; Score 22; DB 25; Length 17646;

Best Local Similarity 100.0%; Pred. No. 1.7;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1133 TAAATATTATTCGAAATCTTTC 1154

9447 TAAATATTATTCGAAATCTTTC 9426

RESULT 14

ABT17023/C

ID ABT17023 standard; DNA; 17646 BP.

AC ABT17023;

03-APR-2003 (first entry)

Human secreted protein-related DNA sequence - SEQ ID No 377.

Human; gene; ds; protein therapy; immediate hypersensitivity disease;

allergic disorder; asthmatic disorder; gene therapy; secreted protein; hay fever; allergic conjunctivitis; allergic rhinitis; binding partner identification; chromosome identification; radiation hybrid mapping; long-range restriction mapping.

Homo sapiens.

WO200277188-A2.

03-OCT-2002.

26-MAR-2002; 2002WO-US09239.

27-MAR-2001; 2001US-278650P.

12-SEP-2001; 2001US-0950082.

12-SEP-2001; 2001US-0950083.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Ruben SM;

WPI; 2003-175010/17.

Use of human secreted proteins and nucleic acids for preparing a diagnostic or pharmaceutical composition for diagnosing or treating allergic or asthmatic disorders, e.g. asthma, hay fever, or allergic conjunctivitis or rhinitis

Disclosure; Page 815-819; 823pp; English.

The invention comprises the amino acid and coding sequences of human secreted proteins. The DNA and protein sequences of the invention are useful for the diagnosis and treatment of allergic disorders, asthmatic disorders and immediate hypersensitivity diseases (e.g. hay fever, allergic conjunctivitis and allergic rhinitis). The proteins of the invention are also useful for identifying a binding partner. The nucleic acids of the invention are also useful for chromosome identification, radiation hybrid mapping or long-range restriction mapping. The present DNA sequence represents a human secreted protein-related DNA sequence.

Sequence 17646 BP; 5370 A; 3099 C; 3561 G; 5616 T; 0 other;

Query Match 1.3%; Score 22; DB 25; Length 17646;

Best Local Similarity 100.0%; Pred. No. 1.7;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1133 TAAATATTATTCGAAATCTTTC 1154

9447 TAAATATTATTCGAAATCTTTC 9426

RESULT 15

ABZ68163/C

ID ABZ68163 standard; DNA; 17646 BP.

AC ABZ68163;

26-MAR-2003 (first entry)

Human secreted protein encoding genomic DNA SEQ ID NO 1686.

Human; secreted protein; nontropic; neuroprotective; cytostatic; virucide; dermatological; immunosuppressive; anti-inflammatory; anti-HIV; vulnery; antibacterial; antiparkinsonian; antisking; antianemic; antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective; antiinflammatory; anti-allergic; antidiabetic; antitumor; anticonvulsant; antifungal; antiparasitic; cardiac; immune disorder; infection; vaccine; cardiovascular disorder; neurological disease; nephrotropic; gene therapy; gene; ds.

Homo sapiens.

WO200277186-A2.

03-OCT-2002.
26-MAR-2002; 2002WO-US09188.
27-MAR-2001; 2001US-278650P.
12-SEP-2001; 2001US-0950082.
12-SEP-2001; 2001US-0950083.
(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Ruben SM;
WPI; 2003-040583/03.
New human secreted proteins encoded by genes contained in cDNA clones
(e.g. HCA19), useful for preventing, treating or diagnosing e.g.
AIDS, multiple sclerosis, herpes virus, leukemia, tick-borne
encephalitis or West Nile fever -
Disclosure; Page 2326-2330; 2423pp; English.
The invention relates to novel human genes (ABZ66891-ABZ68209) and the
encoded secreted proteins (ABP99470-ABP99872) useful for preventing,
treating or ameliorating medical conditions e.g. by protein or gene
therapy. The genes are isolated from a range of human tissues disclosed
in the specification. The nucleic acids, proteins, antibodies and
(ant)agonists are useful in the diagnosis, treatment and prevention of:
(a) cancer, e.g. breast and ovarian cancer and other cancers of the
adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
lung or urogenital; (b) immune disorders e.g. Addison's disease,
allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
arthritis and ulcerative colitis; (c) cardiovascular disorders such as
myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.
cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
bacterial, fungal and parasitic infections.
Sequence 17646 BP; 5370 A; 3099 C; 3561 G; 5616 T; 0 other;
Query Match 1.3%; Score 22; DB 25; Length 17646;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1133 TAAAAATTATTGAAAATCTTTC 1154
9447 TAAAAATTATTGAAAATCTTTC 9426

Search completed: February 2, 2004, 03:13:29
Job time : 496.761 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

M nucleic - nucleic search, using sw model

Run on: February 2, 2004, 00:53:06 ; Search time 3880.05 Seconds
(without alignments)
10504.671 Million cell updates/sec

Title: US-09-938-842A-3729

Perfect score: 1677
Sequence: 1 ggttaaggttttactatg.....ttttctcagctatatatta 1677

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_nam:*
23: em_gss_nus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 271 | 16.2 | 317 | 29 | BZ662927 |
| 2 | 245 | 14.6 | 267 | 29 | AL761108 Arabidops |
| 3 | 164 | 9.8 | 232 | 28 | BH612074 |
| 4 | 164 | 9.8 | 233 | 28 | BH612075 |

| | | | | | |
|----|-----|-----|-----|----|-----------|
| 5 | 150 | 8.9 | 278 | 29 | BZ377781 |
| 6 | 149 | 8.9 | 347 | 9 | AV521636 |
| 7 | 101 | 6.0 | 486 | 9 | AV825375 |
| 8 | 85 | 5.1 | 378 | 10 | BH523004 |
| 9 | 82 | 4.9 | 146 | 28 | BH617024 |
| 10 | 40 | 2.4 | 600 | 29 | CA459772 |
| 11 | 26 | 1.6 | 825 | 28 | BH493162 |
| 12 | 22 | 1.3 | 207 | 29 | AL764163 |
| 13 | 21 | 1.3 | 162 | 28 | AQ055392 |
| 14 | 21 | 1.3 | 178 | 10 | BH933474 |
| 15 | 21 | 1.3 | 258 | 10 | BF703799 |
| 16 | 21 | 1.3 | 269 | 14 | CB477075 |
| 17 | 21 | 1.3 | 310 | 9 | AA224031 |
| 18 | 21 | 1.3 | 318 | 9 | AW480642 |
| 19 | 21 | 1.3 | 329 | 9 | AW358594 |
| 20 | 21 | 1.3 | 330 | 9 | AA224071 |
| 21 | 21 | 1.3 | 349 | 14 | CD478137 |
| 22 | 21 | 1.3 | 365 | 9 | AI466455 |
| 23 | 21 | 1.3 | 391 | 14 | CA940873 |
| 24 | 21 | 1.3 | 406 | 10 | BF449231 |
| 25 | 21 | 1.3 | 413 | 10 | BE207343 |
| 26 | 21 | 1.3 | 424 | 9 | AI035951 |
| 27 | 21 | 1.3 | 440 | 10 | BF707341 |
| 28 | 21 | 1.3 | 440 | 10 | BE255542 |
| 29 | 21 | 1.3 | 445 | 9 | AUI184321 |
| 30 | 21 | 1.3 | 455 | 9 | AUI184443 |
| 31 | 21 | 1.3 | 476 | 10 | BF707342 |
| 32 | 21 | 1.3 | 488 | 13 | BX281427 |
| 33 | 21 | 1.3 | 521 | 10 | BF704377 |
| 34 | 21 | 1.3 | 551 | 9 | AW655154 |
| 35 | 21 | 1.3 | 555 | 10 | BF284594 |
| 36 | 21 | 1.3 | 564 | 12 | BM384056 |
| 37 | 21 | 1.3 | 587 | 12 | BY537127 |
| 38 | 21 | 1.3 | 599 | 10 | BG539984 |
| 39 | 21 | 1.3 | 635 | 10 | BE252939 |
| 40 | 21 | 1.3 | 637 | 10 | BE254600 |
| 41 | 21 | 1.3 | 663 | 28 | B53012 |
| 42 | 21 | 1.3 | 702 | 14 | CB309574 |
| 43 | 21 | 1.3 | 707 | 13 | B0194910 |
| 44 | 21 | 1.3 | 711 | 10 | BE251436 |
| 45 | 21 | 1.3 | 728 | 10 | BH610568 |

ALIGNMENTS

RESULT 1

BZ662927

LOCUS

DEFINITION

Arabidopsis thaliana genomic clone SALK_026421.45-45.x, genomic

survey sequence.

ACCESSION

BZ662927

VERSION

BZ662927.1

KEYWORDS

GSS.

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BZ662927 317 bp DNA linear GSS 31-JAN-2003
SALK_026421.45-45.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_026421.45-45.x, genomic
survey sequence.

BZ662927.1 GI:28177691

GSS.

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eurosids II; Brassicales; Brassicaceae; Arabidopsi

1 (bases 1 to 317)

Alonso,J.M., Leisner,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab

,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.

, Zimmerman,J. and Ecker,J.R.

A Sequence-Indexed Library of Insertion Mutations in the

Arabidopsis Genome

Unpublished

Contact: Joseph R. Ecker

Salk Institute Genomic Analysis Laboratory (SIGAL)

The Salk Institute for Biological Studies

10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

Tel: 858 453 4100 x1752

Fax: 858 558 6379

Email: ecker@salk.edu

This is single pass sequence recovered from the left border of TDNA. This sequence lies within 300 bases of the 5' end of At2g45680.

Class: TDNA tagged.

Location/Qualifiers

1. 317
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_024421.45.45.x"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html"

BASE COUNT 86 a 75 c 51 g 105 t

ORIGIN

Query Match 16.2%; Score 271; DB 29; Length 317;

Best Local Similarity 100.0%; Pred. No. 1.8e-126;

Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2y 1407 GAGCAATGGAAGTATTTGTCACGTGTACAAATCGAATGGTCTTTAAAGCTCATCG 1466

2b 27 GAGCAATGGAAGTATTTGTCACGTGTACAAATCGAATGGTCTTTAAAGCTCATCG 86

2y 1467 AACACATCAGGACCGTTGATTTTCCCGCATCAAAAGCGTTGAATACATTTCTCACTTG 1526

2b 87 AACACATCAGGACCGTTGATTTTCCCGCATCAAAAGCGTTGAATACATTTCTCACTTG 146

2y 1527 TTTTCTGCTCCAT 1586

2b 147 TTTTCTGCTCCAT 206

2y 1587 AACCGCGTTAAACAGTATCTTCCCATCTGATCCGCTTTTACAACTCTCTGCTCATCT 1646

2b 207 AACCGCGTTAAACAGTATCTTCCCATCTGATCCGCTTTTACAACTCTCTGCTCATCT 266

2y 1647 CCACCGTCCGTTTCTCTCAGCTATATTTTA 1677

2b 267 CCACCGTCCGTTTCTCTCAGCTATATTTTA 297

RESULT 2

AL761108

LOCUS

DEFINITION AL761108 267 bp DNA linear GSS 18-JUN-2002

Arabidopsis thaliana T-DNA flanking sequence GK-205F06-014511,

genomic survey sequence.

ACCESSION AL761108.1 GI:21502149

VERSION GSS.

KEYWORDS Arabidopsis thaliana (thale cress)

SOURCE Arabidopsis thaliana

ORGANISM Arabidopsis thaliana (thale cress)

Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi

1 Strizhov,N., Li,Y., Rosso,M., Viehoever,P., Dekker,K., Saedler,H.

and Weisshaar,B.

A pipeline for automated high-throughput generation of FSTs

(flanking sequence tags) from Arabidopsis thaliana T-DNA

transformed lines

unpublished

2

REFERENCE

A new Arabidopsis thaliana T-DNA mutagenised population (GABI-kat)

for flanking sequence tag based reverse genetics

Unpublished

3 (bases 1 to 267)

AUTHORS

Direct Submission

JOURNAL

COMMENT

1. 267

Location/Qualifiers

source

1. 267

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/strain="Columbia 0"

/db_xref="taxon:3702"

/clone="GK-205F06-014511"

/clone_lib="Arabidopsis thaliana T-DNA insertion lines"

/note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector pAC161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the A. thaliana nuclear genome sequences were processed for submission. T-DNA derived sequences were removed"

BASE COUNT 64 a 48 c 68 g 87 t

ORIGIN

Query Match 14.5%; Score 245; DB 29; Length 267;

Best Local Similarity 100.0%; Pred. No. 3.2e-113;

Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 390 AGTTGGTGGCTGGCAAAATACAACTATTCCTCTGTGTGACCCGGAATCTGTAATCG 449

Db 23 AGTTGGTGGCTGGCAAAATACAACTATTCCTCTGTGTGACCCGGAATCTGTAATCG 82

Qy 450 GAAAGGTGGAACCCACTTGGTTTAACTTTTAAGCTTAAAGCTTACTACCGTTTGACC 509

Db 83 GAAAGGTGGAACCCACTTGGTTTAACTTTTAAGCTTAAAGCTTACTACCGTTTGACC 142

Qy 510 GGTTTATAATTTGGTGTGTTTAAATCTTAATCCCGATCCGTTTGTGTTTAACTCTCAAGGCC 569

Db 143 GGTTTATAATTTGGTGTGTTTAAATCTTAATCCCGATCCGTTTGTGTTTAACTCTCAAGGCC 202

Qy 570 ACGTATCGCAATATTTGANTTTTGAGTGGTAGGAAATGTTGGGGTCAATAGTTGG 629

Db 203 ACGTATCGCAATATTTGANTTTTGAGTGGTAGGAAATGTTGGGGTCAATAGTTGG 262

Qy 630 GCCTA 634

Db 263 GCCTA 267

RESULT 3

BH612074/c

LOCUS

DEFINITION SALK_032102 Arabidopsis thaliana TDNA insertion lines Arabidopsis

thaliana genomic clone SALK_032102, genomic survey sequence.

ACCESSION BH612074

VERSION BH612074.1 GI:18059525

KEYWORDS GSS.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

1 (bases 1 to 232)

REFERENCE

A new Arabidopsis thaliana T-DNA mutagenised population (GABI-kat)

for flanking sequence tag based reverse genetics

Unpublished

3 (bases 1 to 267)

Rosso,M., Li,Y., Strizhov,N. and Weisshaar,B.

Submitted (17-JUN-2002) Weisshaar B., Max-Planck-Institut fuer

Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany

This sequence is recovered from the left border of the T-DNA. It

indicates an insertion within the locus defined by clone f418. The

sequences are generated at the MPI for Plant Breeding Research in

the context of the GABI-Kat project. GABI-Kat is part of the German

Plant Genomics program designated 'GABI'. Information on line

availability can be found at:

<http://www.mpiz-koeln.mpg.de/GABI-Kat/>.

Location/Qualifiers

1. 267

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/strain="Columbia 0"

/db_xref="taxon:3702"

/clone="GK-205F06-014511"

/clone_lib="Arabidopsis thaliana T-DNA insertion lines"

/note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector pAC161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the A. thaliana nuclear genome sequences were processed for submission. T-DNA derived sequences were removed"

BASE COUNT 64 a 48 c 68 g 87 t

ORIGIN

Query Match 14.5%; Score 245; DB 29; Length 267;

Best Local Similarity 100.0%; Pred. No. 3.2e-113;

Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 390 AGTTGGTGGCTGGCAAAATACAACTATTCCTCTGTGTGACCCGGAATCTGTAATCG 449

Db 23 AGTTGGTGGCTGGCAAAATACAACTATTCCTCTGTGTGACCCGGAATCTGTAATCG 82

Qy 450 GAAAGGTGGAACCCACTTGGTTTAACTTTTAAGCTTAAAGCTTACTACCGTTTGACC 509

Db 83 GAAAGGTGGAACCCACTTGGTTTAACTTTTAAGCTTAAAGCTTACTACCGTTTGACC 142

Qy 510 GGTTTATAATTTGGTGTGTTTAAATCTTAATCCCGATCCGTTTGTGTTTAACTCTCAAGGCC 569

Db 143 GGTTTATAATTTGGTGTGTTTAAATCTTAATCCCGATCCGTTTGTGTTTAACTCTCAAGGCC 202

Qy 570 ACGTATCGCAATATTTGANTTTTGAGTGGTAGGAAATGTTGGGGTCAATAGTTGG 629

Db 203 ACGTATCGCAATATTTGANTTTTGAGTGGTAGGAAATGTTGGGGTCAATAGTTGG 262

Qy 630 GCCTA 634

Db 263 GCCTA 267

RESULT 3

BH612074/c

LOCUS

DEFINITION SALK_032102 Arabidopsis thaliana TDNA insertion lines Arabidopsis

thaliana genomic clone SALK_032102, genomic survey sequence.

ACCESSION BH612074

VERSION BH612074.1 GI:18059525

KEYWORDS GSS.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

1 (bases 1 to 232)

REFERENCE

A new Arabidopsis thaliana T-DNA mutagenised population (GABI-kat)

for flanking sequence tag based reverse genetics

Unpublished

3 (bases 1 to 267)

TITLE A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
JOURNAL Unpublished
COMMENT Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel.: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of TDNA.

FEATURES
source
Location/Qualifiers
1..232
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_032104"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html"

BASE COUNT 59 a 45 c 46 g 82 t
ORIGIN
Query Match 9.8%; Score 164; DB 28; Length 232;
Best Local Similarity 100.0%; Pred. No. 6.3e-72;
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 183 AGCTTCTCTATGCTTAAGAAGATGACCGATACGATTAACACAGCATCAATTAAGATTA 242
D 188 AGCTTCTCTATGCTTAAGAAGATGACCGATACGATTAACACAGCATCAATTAAGATTA 129
243 AATGTTTGTGAAGAAATCACTACCTTATTATGGAATTTGTGGTTAGTGAAGATTA 302
D 128 AATGTTTGTGAAGAAATCACTACCTTATTATGGAATTTGTGGTTAGTGAAGATTA 69
Y 303 AAACATCGGAATCCAAACCTCAATTTACCAATCAGCCCAAT 346
D 68 AAACATCGGAATCCAAACCTCAATTTACCAATCAGCCCAAT 25

RESULT 4
H612075/c
OCUS
DEFINITION SALK_032104 Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_032104, genomic survey sequence.
CCESSION B612075
ESSION B612075.1 GI:18059527
EYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 233)
Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrihab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J., and Ecker, J.R.
A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
Unpublished
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel.: 858 453 4100 x1752
Fax: 858 558 6379

Email: ecker@salk.edu
This is single pass sequence recovered from the left border of TDNA.
Class: TDNA tagged.

FEATURES
source
Location/Qualifiers
1..233
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_032104"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html"

BASE COUNT 58 a 45 c 47 g 83 t
ORIGIN
Query Match 9.8%; Score 164; DB 28; Length 233;
Best Local Similarity 100.0%; Pred. No. 6.3e-72;
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 183 AGCTTCTCTATGCTTAAGAAGATGACCGATACGATTAACACAGCATCAATTAAGATTA 242
Db 188 AGCTTCTCTATGCTTAAGAAGATGACCGATACGATTAACACAGCATCAATTAAGATTA 129
QY 243 AATGTTTGTGAAGAAATCACTACCTTATTATGGAATTTGTGGTTAGTGAAGATTA 302
Db 128 AATGTTTGTGAAGAAATCACTACCTTATTATGGAATTTGTGGTTAGTGAAGATTA 69
QY 303 AAACATCGGAATCCAAACCTCAATTTACCAATCAGCCCAAT 346
Db 68 AAACATCGGAATCCAAACCTCAATTTACCAATCAGCCCAAT 25

RESULT 5
BZ377781
LOCUS 278 bp DNA linear GSS-26-NOV-2002
DEFINITION SALK_106185.39.30.x Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_106185.39.30.x, genomic survey sequence.
ACCESSION BZ377781
VERSION BZ377781.1 GI:25467878
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 278)
Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrihab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J., and Ecker, J.R.
A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
Unpublished
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel.: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of TDNA. This sequence lies within 300 bases of the 5' end of At2g45680.

FEATURES
source
Location/Qualifiers
1..278
/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone="SALK 106185.39.30.x"
 /clone_lib="Arabidopsis thaliana TDNA insertion lines"
 /note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html"

BASE COUNT 72 a 63 c 44 g 93 t
 ORIGIN

Query Match 8.9%; Score 150; DB 29; Length 278;
 Best Local Similarity 100.0%; Pred. No. 8.7e-65;
 Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1528 TTTCCTGCTCTATATATATATCTGACGAGTCACATTAGTAATCTCTTGACGTGA 1587
 Db 109 TTTCCTGCTCTATATATATATCTGACGAGTCACATTAGTAATCTCTTGACGTGA 168
 QY 1588 ACGCCGTTAAACGATCTTTCCCATTTGTATCCGCTTTTAAACAACCTCTGCTCATCTC 1647
 Db 169 ACGCCGTTAAACGATCTTTCCCATTTGTATCCGCTTTTAAACAACCTCTGCTCATCTC 228
 QY 1648 CACGTCGCTTTCTTCACGATATATTTA 1677
 Db 229 CACGTCGCTTTCTTCACGATATATTTA 258

RESULT 6
 AV521636/c
 LOCUS AV521636 Arabidopsis thaliana aboveground organs two to six-week
 DEFINITION Old Arabidopsis thaliana cDNA clone AP263b12F 3', mRNA sequence.
 ACCESSION AV521636
 VERSION AV521636.1 GI:8681163
 KEYWORDS EST.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; euroids II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE 1 (bases 1 to 347)
 AUTHORS Asamizu E., Nakamura Y., Sato S. and Tabata S.
 TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries

JOURNAL DNA Res. 7, 175-180 (2000)
 MEDLINE 20363093
 PUBMED 10907847

COMMENT Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: <http://www.kazusa.or.jp/en/plant/>.

FEATURES
 Location/Qualifiers
 1..347
 /organism="Arabidopsis thaliana"
 /mol_type="cDNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="AP263b12F"
 /tissue_type="aboveground organs"
 /dev_stage="two to six-week old"
 /clone_lib="Arabidopsis thaliana aboveground organs two to six-week old"
 /note="vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 102 a 77 c 54 g 114 t
 ORIGIN

Query Match 8.9%; Score 149; DB 9; Length 347;
 Best Local Similarity 100.0%; Pred. No. 2.8e-64;
 Matches 149; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTAAAGCGTTTACTTATGTTATATGCAACGGAGATATGCAATTTGGATGC 60
 Db 241 GGTAAAGCGTTTACTTATGTTATATGCAACGGAGATATGCAATTTGGATGC 182
 QY 61 TTTTTCAGATCATCAAGGCTCCCTACAGATTCTTCTAGGGAATGTTTCAGGCTTTTGTTA 120
 Db 181 TTTTTCAGATCATCAAGGCTCCCTACAGATTCTTCTAGGGAATGTTTCAGGCTTTTGTTA 122
 QY 121 GAAATTGCTTTATTGCAACAGGTAGAGA 149
 Db 121 GAAATTGCTTTATTGCAACAGGTAGAGA 93

RESULT 7
 AV825375
 LOCUS AV825375 486 bp mRNA linear EST 01-APR-2002
 DEFINITION AV825375 RAF17 Arabidopsis thaliana cDNA clone RAFL07-08-P04 5', mRNA sequence.

ACCESSION AV825375
 VERSION AV825375.1 GI:19867435
 KEYWORDS EST.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; euroids II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE 1 (bases 1 to 486)
 AUTHORS Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M., Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y., Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y. and Shinozaki, K.

TITLE Large scale analysis of Arabidopsis full-length cDNA (2002b)
 JOURNAL Unpublished
 COMMENT Contact: Motoaki Seki
 Plant Functional Genomics Research Group
 RIKEN Genomic Sciences Center
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: 81-298-36-4359
 Fax: 81-298-36-9060
 Email: mesaki@rtc.riken.go.jp

An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998). cDNA cleaved with BamHI and XhoI was ligated to modified Lambda FLC-1 vector (Carninci et al., submitted for publication) digested with BamHI and SalI. This clone is in a modified pBluescript vector. Please visit our web site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further details.

FEATURES
 Location/Qualifiers
 1..486
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /clone="RAFL07-08-P04"
 /dev_stage="rosette plants"
 /lab_host="DH10B"
 /clone_lib="RAFL7"
 /note="Site_1: BamHI; Site_2: SalI; subjected to cold-treated (1, 2, 5, 10, 24 hr.)"

BASE COUNT 131 a 118 c 117 g 117 t
 ORIGIN

Query Match 6.0%; Score 101; DB 9; Length 486;
 Best Local Similarity 100.0%; Pred. No. 8.4e-40;
 Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1577 TTGACGTGTAAACCGCTTAAACGATTTCTTCCCATTTGATCCGTTTAAACATCTC 1636
 Db 1 TTGACGTGTAAACCGCTTAAACGATTTCTTCCCATTTGATCCGTTTAAACATCTC 60

```

Y 1637 GTTCGTCATCTCCACCGTCGGTTTCTCTCAGCTATATTTA 1677
b 61 GTTCGTCATCTCCACCGTCGGTTTCTCTCAGCTATATTTA 101

RESULT 8
BES23004
LOCUS
DEFINITION
M31B6STM Arabidopsis developing seed Arabidopsis thaliana cDNA
clone M31B6 5', mRNA sequence.
ACCESSION
BES23004
VERSION
BES23004
KEYWORDS
EST.
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 1 to 378)
AUTHORS
White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de
Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.
TITLE
A new set of Arabidopsis expressed sequence tags from developing
seeds. The metabolic pathway from carbohydrates to seed oil
JOURNAL
Plant Physiol. 124 (4), 1582-1594 (2000)
MEDLINE
20567808
PUBMED
11115876
COMMENT
Contact: Benning, C
Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824
, USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Michigan State University DNA Sequencing Facility Arabidopsis
Biological Resource Center, The Ohio State University, 309 Botany &
Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:
6142920603 TEL: 6142929371.

FEATURES
source
1..378
Location/Qualifiers
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="M31B6"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/clone_lib="Arabidopsis developing seed"
/note="Organ: Developing seed; Vector: pBluescript SK-;
Site 1: EcoRI; Site 2: XhoII"
BASE COUNT 108 a 97 c 81 g 92 t
ORIGIN
1..378
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="M31B6"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/clone_lib="Arabidopsis developing seed"
/note="Organ: Developing seed; Vector: pBluescript SK-;
Site 1: EcoRI; Site 2: XhoII"
Query Match 5.1%; Score 85; DB 10; Length 378;
Best Local Similarity 100.0%; Pred.No. 1.2e-31;
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1593 GTTAAACGATCTTTCCTCCATGTCATCGCTTTTAAACAACCTCGTCGTCATCTCCACCG 1652
b 18 GTTAAACGATCTTTCCTCCATGTCATCGCTTTTAAACAACCTCGTCGTCATCTCCACCG 77

RESULT 9
BH617024
LOCUS
DEFINITION
SALK 035853 Arabidopsis thaliana DNA insertion lines Arabidopsis
thaliana genomic clone SALK_035853, genomic survey sequence.
ACCESSION
BH617024

```

```

VERSION
KEYWORDS
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 1 to 146)
AUTHORS
Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.
,Zimmerman,J. and Ecker,J.R.
TITLE
A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome
JOURNAL
Unpublished
COMMENT
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA. This sequence lies within 300 bases of the 5' end of
At2G45680.
Class: TDNA tagged.
FEATURES
Location/Qualifiers
1..146
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_035853"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna\_protocols.html"
BASE COUNT 33 a 44 c 27 g 42 t
ORIGIN
4..9%; Score 82; DB 28; Length 146;
Best Local Similarity 100.0%; Pred.No. 4e-30;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1596 AAAACGATCTTTCCTCCATGTCATCGCTTTTAAACAACCTCGTCGTCATCTCCACCGTCC 1655
Db 1 AAAACGATCTTTCCTCCATGTCATCGCTTTTAAACAACCTCGTCGTCATCTCCACCGTCC 60

QY 1656 GTTTTCTCTCAGCTATATTTA 1677
Db 61 GTTTTCTCTCAGCTATATTTA 82

Query Match 4.9%; Score 82; DB 28; Length 146;
Best Local Similarity 100.0%; Pred.No. 4e-30;
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 10
CC459772/c
LOCUS
DEFINITION
CC459772 600 bp DNA linear GSS 03-JUN-2003
SALK 133376.34.05.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_133376.34.05.x, genomic
survey sequence.
ACCESSION
CC459772
VERSION
CC459772.1 GI:31343739
KEYWORDS
GSS.
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 1 to 600)
AUTHORS
Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.
,Zimmerman,J. and Ecker,J.R.

```

TITLE A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
 JOURNAL Unpublished
 COMMENT Contact: Joseph R. Ecker
 Salk Institute Genomic Analysis Laboratory (SIGAL)
 The Salk Institute for Biological Studies
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
 Tel: 858 453 4100 x1752
 Fax: 858 558 6379
 Email: ecker@salk.edu
 This is single pass sequence recovered from the left border of TDNA.

Class: TDNA tagged.

Location/Qualifiers

FEATURES

source

1. .600

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/strain="Columbia 0"

/db_xref="taxon:3702"

/clone="SALK_133376.34.05.x"

/notes="Arabidopsis thaliana TDNA insertion lines"
 /clone_lib="Arabidopsis thaliana TDNA insertion lines"
 /notes="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html"

160 a 130 c 102 g 166 t 42 others

Query Match 2.4%; Score 40; DB 29; Length 600;

Best Local Similarity 100.0%; Pred. No. 1.1e-08;

Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 1057 TAATGGGTCCACCCCAAGTATGGCTTACAGCTTTTC 1096

475 TAATGGGTCCACCCCAAGTATGGGTACAGCTTTTC 436

RESULT 11

BH493162/c

LOCUS

DEFINITION BH493162 825 bp DNA linear GSS 13-DEC-2001
 BOGOH91TR BOGO Brassica oleracea genomic clone BOGOH91, genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Brassica oleracea
 Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 825)

Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.

Whole genome shotgun sequencing of Brassica oleracea

Unpublished

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

FEATURES

source

1. .825

/organism="Brassica oleracea"

/mol_type="genomic DNA"

/strain="TO1000DH3"

/db_xref="taxon:3712"

/clone="BOGOH91"

/clone_lib="BOGO"

/note="Vector: pROSI; Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into pROSI using BstXI linkers"

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 26; Conservative

0; Mismatches

0; Indels

0; Gaps

0;

QY 1058 AATGGGTCCACCCCAAGTATGGC 1083

Db 91 AATGGGTCCACCCCAAGTATGGC 66

RESULT 12

AL764163

LOCUS

DEFINITION

Arabidopsis thaliana T-DNA flanking sequence GK-122B06-012551,

genomic survey sequence.

ACCESSION

AL764163

VERSION

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Eukaryota; Viridiplantae;

Streptophyta; Embryophyta;

Tracheophyta;

Spermatophyta; Magnoliophyta;

eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales;

Brassicaceae; Arabidopsis.

REFERENCE

AUTHORS

Strizhov, N., Li, Y., Rosso, M., Viehoever, P., Dekker, K., Saedler, H.

and Weisshaar, B.

A pipeline for automated high-throughput generation of FSTs

(flanking sequence tags) from Arabidopsis thaliana T-DNA

transformed lines

Unpublished

JOURNAL

REFERENCE

AUTHORS

Li, Y., Rosso, M., Strizhov, N. and Weisshaar, B.

Direct Submission

Submitted (17-JUN-2002)

Weisshaar, B., Max-Planck-Institut fuer

Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany

This sequence is recovered from the left border of the T-DNA. It

indicates an insertion within the locus defined by clone F4H6. The

sequences are generated at the MPI for Plant Breeding Research in

the context of the GABI-Kat project. GABI-Kat is part of the German

Plant Genomics program designated 'GABI'. Information on line

availability can be found at:

<http://www.mpiz-koein.mpg.de/GABI-Kat/>.

Location/Qualifiers

FEATURES

source

1. .207

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/strain="Columbia 0"

/db_xref="taxon:3702"

/clone="GK-122B06-012551"

/clone_lib="Arabidopsis thaliana T-DNA insertion lines"

/note="PCR was performed on DNA from Arabidopsis thaliana

plants (T1) which were transformed with the T-DNA from

vector PAC161. The lines contain one or more T-DNA

insertions. The DNA fragment(s) resulting from the PCR

were directly sequenced to determine the genomic sequence

flanking the insertion. Sequences displaying significant

similarity to the A. thaliana nuclear genome sequence were

processed for submission. T-DNA derived sequences were

removed"

BASE COUNT 88 a 35 c 33 g 51 t

ORIGIN

Query Match

Best Local Similarity

1.3%; Score 22; DB 29;

Length 207;

Matches 100.0%; Pred. No. 16;

```

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

>y 1308 ATTACTAAACATTAAGAAAA 1329
|||||
>b 141 ATTACTAAACATTAAGAAAA 162
|||||

RESULT 13
LOCUS Q055392
DEFINITION CIT-HSP-2344P6, TR CIT-HSP Homo sapiens genomic clone 2344P6,
genomic survey sequence.
ACCESSION Q055392
VERSION Q055392.1 GI:3351998
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 162)
Adams, M.D., Rounsley, S.D., Zhao, S., Field, C.E., Bass, S., Linher, K.,
Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H.,
Simon, M., and Venter, J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
JOURNAL Unpublished
COMMENT Other GSSs: CIT-HSP-2344P6.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.
FEATURES
source
1..162
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="2344P6"
/sex="Male"
/cell_type="sperm"
/clone_lib="CIT-HSP"
/note="Vector: pBelosAC11; Site_1: HindIII; Site_2:
HindIII"
BASE COUNT 56 a 36 c 22 g 48 t
ORIGIN
Query Match 1.3%; Score 21; DB 28; Length 162;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

>y 1093 TTTCATTAATTAAGTAAA 1113
|||||
>b 103 TTTCATTAATTAAGTAAA 123
|||||

RESULT 14
LOCUS BE933474/c
DEFINITION RC4-HT0887-250800-012-d07 HT0887 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE933474
VERSION BE933474.1 GI:10459550
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
178 bp mRNA linear EST 02-OCT-2000
RC4-HT0887-250800-012-d07 HT0887 Homo sapiens cDNA, mRNA sequence.
1 (bases 1 to 178)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, P.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J., and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
MEDLINE 10737800
PUBMED
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=RC4-HT0887-250
800-012-d07&t3=2000-08-25&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 16
High quality sequence stop: 154.
FEATURES
source
1..178
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="HT0887"
/note="Organ: head neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 40 a 54 c 45 g 39 t
ORIGIN
Query Match 1.3%; Score 21; DB 10; Length 178;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

>y 416 TTATCCCTCTGTGTGACCC 436
|||||
>b 43 TTATCCCTCTGTGTGACCC 23
|||||

RESULT 15
LOCUS BF703799
DEFINITION MI-P-E4-abg-h-11-1-UM.s1 MI-P-E4 Sus scrofa cDNA clone
MI-P-E4-abg-h-11-1-UM 3', mRNA sequence.
ACCESSION BF703799
VERSION BF703799.1 GI:11989207
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 258)
Bonaldi, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL 9704477
MEDLINE 8889548
PUBMED
Contact: Tuggle CK

```

```

REFERENCE
AUTHORS
1 (bases 1 to 178)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, P.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J., and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
MEDLINE 10737800
PUBMED
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=RC4-HT0887-250
800-012-d07&t3=2000-08-25&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 16
High quality sequence stop: 154.
FEATURES
source
1..178
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="HT0887"
/note="Organ: head neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 40 a 54 c 45 g 39 t
ORIGIN
Query Match 1.3%; Score 21; DB 10; Length 178;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

>y 416 TTATCCCTCTGTGTGACCC 436
|||||
>b 43 TTATCCCTCTGTGTGACCC 23
|||||

RESULT 15
LOCUS BF703799
DEFINITION MI-P-E4-abg-h-11-1-UM.s1 MI-P-E4 Sus scrofa cDNA clone
MI-P-E4-abg-h-11-1-UM 3', mRNA sequence.
ACCESSION BF703799
VERSION BF703799.1 GI:11989207
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 258)
Bonaldi, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL 9704477
MEDLINE 8889548
PUBMED
Contact: Tuggle CK

```


Molecular Genetics Laboratory, Department of Animal Science
Iowa State University
201 Kildee Hall, Ames, IA 50011-3150, USA
Tel: 5152944252
Fax: 5152942401
Email: cktuggle@iastate.edu

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
non-normalized embryo at gestational day 14 library cDNA library
Preparation: RJ Woods, JA Green, RS Prather S142 Animal Science
Research Center, Department of Animal Science, University of
Missouri-Columbia, 65211 Clone distribution: clones will be
available through Research Genetics (www.resgen.com)

Seq primer: M13 Forward

POLYA=Yes.

FEATURES

source

Location/Qualifiers

1..258
/organism="Sus scrofa"
/mol_type="mRNA"
/strain="crossbreed"
/db_xref="taxon:9823"
/clone="MI-P-E4-abg-h-11-1-UM"
/clone_lib="MI-P-E4"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: EcoRI; The MI-P-E4
library is derived from embryo at gestational day 14. For
a detailed description of the library from which this
clone was derived, please visit our web site at
<http://pigest.genome.iastate.edu/>.
TAG LIB=MI-P-E4
TAG TISSUE=embryo at gestational day 14
TAG_SEQ=ACTCAC"

BASE COUNT 79 a 52 g 96 t

ORIGIN

Query Match 1.3%; Score 21; DB 10; Length 258;
Best Local Similarity 100.0%; Pred.No.51;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 284 TGCGTTAGTGAAGTAAAA 304

Db 150 TGCGTTAGTGAAGTAAAA 170

Search completed: February 2, 2004, 05:01:24
Job time : 3888.05 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

M nucleic - nucleic search, using sw model

un on: February 2, 2004, 02:56:51 ; Search time 111.678 Seconds
(without alignments)
6627.983 Million cell updates/sec

itle: US-09-938-842A-3729

erfect score: 1677
equence: 1 ggtaagcgtttactatg.....ttctctcagctatattta 1677

oring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

earched: 569978 seqs, 220691566 residues

ord size : 0

otal number of hits satisfying chosen parameters: 1139956

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Listing first 45 summaries

atabase : Issued Patents NA:
1: /cgm2_6/ptodata/2/ina/5A_COMB.seq:
2: /cgm2_6/ptodata/2/ina/5B_COMB.seq:
3: /cgm2_6/ptodata/2/ina/6A_COMB.seq:
4: /cgm2_6/ptodata/2/ina/6B_COMB.seq:
5: /cgm2_6/ptodata/2/ina/PTCTUS_COMB.seq:
6: /cgm2_6/ptodata/2/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|---------|----|---------------------|
| 1 | 18 | 1.1 | 832 | 1 | US-08-340-539A-6 |
| 2 | 18 | 1.1 | 832 | 2 | US-08-461-592B-6 |
| 3 | 18 | 1.1 | 876 | 4 | US-09-198-119C-48 |
| 4 | 18 | 1.1 | 877 | 4 | US-09-198-119C-72 |
| 5 | 18 | 1.1 | 887 | 4 | US-09-198-119C-68 |
| 6 | 18 | 1.1 | 1028 | 3 | US-09-249-180-1 |
| 7 | 18 | 1.1 | 1132 | 4 | US-09-198-119C-56 |
| 8 | 18 | 1.1 | 1163 | 3 | US-09-249-180-5 |
| 9 | 18 | 1.1 | 2148 | 4 | US-09-328-352-2704 |
| 10 | 18 | 1.1 | 580073 | 4 | US-08-545-528D-1 |
| 11 | 18 | 1.1 | 4403765 | 3 | US-09-103-840A-2 |
| 12 | 18 | 1.1 | 4411529 | 3 | US-09-103-840A-1 |
| 13 | 17 | 1.0 | 28 | 1 | US-08-120-827-71 |
| 14 | 17 | 1.0 | 28 | 1 | US-08-478-675-71 |
| 15 | 17 | 1.0 | 210 | 4 | US-09-134-001C-2588 |
| 16 | 17 | 1.0 | 241 | 4 | US-09-389-681-400 |
| 17 | 17 | 1.0 | 241 | 4 | US-09-620-405B-400 |
| 18 | 17 | 1.0 | 241 | 4 | US-09-433-826B-400 |
| 19 | 17 | 1.0 | 241 | 4 | US-09-604-287A-400 |
| 20 | 17 | 1.0 | 267 | 4 | US-09-313-294A-1630 |
| 21 | 17 | 1.0 | 326 | 1 | US-08-525-596B-7 |
| 22 | 17 | 1.0 | 326 | 3 | US-09-177-860A-7 |
| 23 | 17 | 1.0 | 326 | 4 | US-09-378-238-7 |
| 24 | 17 | 1.0 | 326 | 4 | US-09-451-501-7 |
| 25 | 17 | 1.0 | 326 | 4 | US-09-629-938-7 |
| 26 | 17 | 1.0 | 835 | 4 | US-09-171-203-42 |
| 27 | 17 | 1.0 | 938 | 4 | US-09-843-472-2 |
| C | 27 | | | | |

28 17 1.0 1128 4 US-09-252-149B-1 Sequence 1, Appli
29 17 1.0 1128 4 US-09-451-501-20 Sequence 20, Appli
30 17 1.0 1128 4 US-09-451-501-26 Sequence 26, Appli
31 17 1.0 1196 3 US-08-891-789B-1 Sequence 1, Appli
32 17 1.0 1240 3 US-08-891-789B-3 Sequence 3, Appli
33 17 1.0 1326 3 US-09-100-391-1 Sequence 1, Appli
34 17 1.0 1326 4 US-09-616-614-1 Sequence 1, Appli
35 17 1.0 1746 4 US-09-107-532A-1731 Sequence 1731, Ap
36 17 1.0 1748 4 US-09-620-312D-60 Sequence 60, Appli
37 17 1.0 1952 2 US-08-481-337A-7 Sequence 7, Appli
38 17 1.0 1952 3 US-09-382-256-17 Sequence 17, Appli
39 17 1.0 1952 3 US-09-395-115-17 Sequence 17, Appli
40 17 1.0 1952 4 US-08-436-285-17 Sequence 17, Appli
41 17 1.0 1952 4 US-09-679-187-17 Sequence 17, Appli
42 17 1.0 1952 5 PCT-US95-05467-7 Sequence 7, Appli
43 17 1.0 1990 4 US-08-961-527-232 Sequence 232, App
44 17 1.0 2076 3 US-08-123-934A-3 Sequence 3, Appli
45 17 1.0 2076 5 PCT-US94-10080-3 Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-340-539A-6
Sequence 6, Application US/08340539A
Patent No. 5880025
GENERAL INFORMATION:
APPLICANT: Tedder, Thomas F.
APPLICANT: Kansas, Geoffrey S.
TITLE OF INVENTION: CHIMERIC SELECTINS AS SIMULTANEOUS
BLOCKING AGENTS FOR COMPONENT SELECTIN FUNCTION
TITLE OF INVENTION: CHIMERIC SELECTINS AS SIMULTANEOUS
BLOCKING AGENTS FOR COMPONENT SELECTIN FUNCTION
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSER: FISH & NEAVE
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/340,539A
FILING DATE: 16-NOV-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/008,459
FILING DATE: 25-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Gumison, Jane
REGISTRATION NUMBER: 38,479
REFERENCE/DOCKET NUMBER: CG-104 CON
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-596-9000
TELEFAX: 212-596-9090
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 832 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-340-539A-6

Query Match 1.1%; Score 18; DB 1; Length 832;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 TCTATCTCTAAAGAAATG 206

db 86 TCTATGCTAAAGAAATG 103

RESULT 2
US-08-461-592B-6
Sequence 6, Application US/08461592B
Patent No. 5834425
GENERAL INFORMATION:
APPLICANT: Tedder, Thomas F.
APPLICANT: Kansas, Geoffrey S.
TITLE OF INVENTION: CHIMERIC SELECTINS AS SIMULTANEOUS
BLOCKING AGENTS FOR COMPONENT SELECTIN FUNCTION
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Weingarten, Schurigin, Gagnebin & Hayes
STREET: Ten Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,592B
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/340,539
FILING DATE: 16-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/008,459
FILING DATE: 25-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: James F. Haley, Jr.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: CG-104
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 596-9000
TELEFAX: (212) 596-9090
TELEX: 14-8367
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 832 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEetical: NO
ANTI-SENSE: NO
US-08-461-592B-6

Query Match 1.1%; Score 18; DB 2; Length 832;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 189 TCTATGCTAAAGAAATG 206
DB 86 TCTATGCTAAAGAAATG 103

RESULT 3
US-09-198-119C-48/c
Sequence 48, Application US/09198119C
Patent No. 6417428
GENERAL INFORMATION:
APPLICANT: Thomashow, Michael
APPLICANT: Stockinger, Eric
APPLICANT: Jaglo-Ottosen, Kirsten
APPLICANT: Gilmour, Sarah

APPLICANT: Zarka, Daniel
APPLICANT: Jiang, Cai-Zhong
TITLE OF INVENTION: Plant Having Altered Environmental Stress Tolerance
FILE REFERENCE: 19117.713 Seq List
CURRENT APPLICATION NUMBER: US/09/198,119C
CURRENT FILING DATE: 1998-11-23
PRIOR APPLICATION NUMBER: US 08/706,270
PRIOR FILING DATE: 1998-09-04
PRIOR APPLICATION NUMBER: US 09/018,233
PRIOR FILING DATE: 1998-02-03
PRIOR APPLICATION NUMBER: US 09/017,816
PRIOR FILING DATE: 1998-02-03
PRIOR APPLICATION NUMBER: US 09/018,235
PRIOR FILING DATE: 1998-02-03
PRIOR APPLICATION NUMBER: US 09/017,575
PRIOR FILING DATE: 1998-02-03
PRIOR APPLICATION NUMBER: US 09/018,227
PRIOR FILING DATE: 1998-02-03
PRIOR APPLICATION NUMBER: US 09/018,234
PRIOR FILING DATE: 1998-02-03
NUMBER OF SEQ ID NOS: 95
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 48
LENGTH: 876
TYPE: DNA
ORGANISM: Brassica napus
FEATURE:
OTHER INFORMATION: bnCBF2 gene
US-09-198-119C-48

Query Match 1.1%; Score 18; DB 4; Length 876;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1364 CCCATTGTACAGATGCT 1381
DB 815 CCCATTGTACAGATGCT 798

RESULT 4
US-09-198-119C-72/c
Sequence 72, Application US/09198119C
Patent No. 6417428
GENERAL INFORMATION:
APPLICANT: Thomashow, Michael
APPLICANT: Stockinger, Eric
APPLICANT: Jaglo-Ottosen, Kirsten
APPLICANT: Gilmour, Sarah
APPLICANT: Zarka, Daniel
APPLICANT: Jiang, Cai-Zhong
TITLE OF INVENTION: Plant Having Altered Environmental Stress Tolerance
FILE REFERENCE: 19117.713 Seq List
CURRENT APPLICATION NUMBER: US/09/198,119C
CURRENT FILING DATE: 1998-11-23
PRIOR APPLICATION NUMBER: US 08/706,270
PRIOR FILING DATE: 1998-09-04
PRIOR APPLICATION NUMBER: US 09/018,233
PRIOR FILING DATE: 1998-02-03
PRIOR APPLICATION NUMBER: US 09/017,816
PRIOR FILING DATE: 1998-02-03
PRIOR APPLICATION NUMBER: US 09/018,235
PRIOR FILING DATE: 1998-02-03
PRIOR APPLICATION NUMBER: US 09/017,575
PRIOR FILING DATE: 1998-02-03
PRIOR APPLICATION NUMBER: US 09/018,227
PRIOR FILING DATE: 1998-02-03
PRIOR APPLICATION NUMBER: US 09/018,234
PRIOR FILING DATE: 1998-02-03
NUMBER OF SEQ ID NOS: 95
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 72
LENGTH: 877
TYPE: DNA

ORGANISM: Brassica oleracea
FEATURE:
OTHER INFORMATION: boCBP5 gene
IS-09-198-119C-72

Query Match 1.1%; Score 18; DB 4; Length 877;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1364 CCCATTGTACAGATGGT 1381
|||||
Db 815 CCCATTGTACAGATGGT 798
|||||

RESULT 5

US-09-198-119C-68/c
Sequence 68, Application US/09198119C
Patent No. 6417428

GENERAL INFORMATION:

APPLICANT: Thomashow, Michael
APPLICANT: Stockinger, Eric
APPLICANT: Jaglo-Ottosen, Kirsten
APPLICANT: Gilmore, Sarah
APPLICANT: Zarka, Daniel
APPLICANT: Jiang, Cai-Zhong

TITLE OF INVENTION: Plant Having Altered Environmental Stress Tolerance

FILE REFERENCE: 1911.713 Seq List

CURRENT APPLICATION NUMBER: US/09198,119C

PRIOR FILING DATE: 1998-11-23

PRIOR APPLICATION NUMBER: US 08/706,270

PRIOR FILING DATE: 1998-09-04

PRIOR APPLICATION NUMBER: US 09/018,233

PRIOR FILING DATE: 1998-02-03

PRIOR APPLICATION NUMBER: US 09/017,816

PRIOR FILING DATE: 1998-02-03

PRIOR APPLICATION NUMBER: US 09/018,235

PRIOR FILING DATE: 1998-02-03

PRIOR APPLICATION NUMBER: US 09/017,575

PRIOR FILING DATE: 1998-02-03

PRIOR APPLICATION NUMBER: US 09/018,227

PRIOR FILING DATE: 1998-02-03

PRIOR APPLICATION NUMBER: US 09/018,234

PRIOR FILING DATE: 1998-02-03

NUMBER OF SEQ ID NOS: 95

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 68

LENGTH: 887

TYPE: DNA

ORGANISM: Brassica oleracea

FEATURE:

OTHER INFORMATION: boCBP3 gene

US-09-198-119C-68

Query Match 1.1%; Score 18; DB 4; Length 887;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1364 CCCATTGTACAGATGGT 1381
|||||
Db 822 CCCATTGTACAGATGGT 805
|||||

RESULT 6

US-09-249-180-1/c

Sequence 1, Application US/09249180

Patent No. 6268548

GENERAL INFORMATION:

APPLICANT: Elthon, Thomas E

APPLICANT: Lund, Adrian A

APPLICANT: Bhattacharjee, Dinakar

APPLICANT: Rhoads, David M.

TITLE OF INVENTION: Isolation and Characterization of Heat Shock Protein

FILE REFERENCE: UNV92819

CURRENT APPLICATION NUMBER: US/09/249,180
CURRENT FILING DATE: 1999-02-12
EARLIER APPLICATION NUMBER: 60/076/014
EARLIER FILING DATE: 1998-02-26
NUMBER OF SEQ ID NOS: 7
SOFTWARE: Created in PatentIn Ver. 2.0, Edited in WordPerfect 6.1
SEQ ID NO 1
LENGTH: 1028
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: CDS
LOCATION: (79)..(735)
FEATURE:
NAME/KEY: Gene
LOCATION: (1)..(1028)
FEATURE:
NAME/KEY: misc feature
LOCATION: (424)..(693)
FEATURE:
NAME/KEY: exon
LOCATION: (1)..(328)
FEATURE:
NAME/KEY: exon
LOCATION: (329)..(1028)
FEATURE:
NAME/KEY: Poly A site
LOCATION: (1028)..(1028)
FEATURE:
NAME/KEY: mRNA
LOCATION: (1)..(1028)
FEATURE:
NAME/KEY: source
LOCATION: (1)..(1028)
FEATURE:
NAME/KEY: transit peptide
LOCATION: (79)..(213)
FEATURE:
NAME/KEY: 3'UTR
LOCATION: (736)..(1028)
FEATURE:
NAME/KEY: 5'UTR
LOCATION: (1)..(78)
PUBLICATION INFORMATION:
AUTHORS: Lund, Adrian A.
AUTHORS: Blum, Paul H.
AUTHORS: Bhattacharjee, Dinakar
AUTHORS: Elthon, Thomas E.
TITLE: Heat-Stress Response of Maize Mitochondria
JOURNAL: Plant Physiol.
VOLUME: 116
PAGES: 1097-1110
DATE: 1998-03-00
US-09-249-180-1

Query Match 1.1%; Score 18; DB 3; Length 1028;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1391 ATACTGATAGAGATAGA 1408
|||||
Db 880 ATACTGATAGAGATAGA 863
|||||

RESULT 7

US-09-198-119C-56/c

Sequence 56, Application US/09198119C

Patent No. 6417428

GENERAL INFORMATION:

APPLICANT: Thomashow, Michael

APPLICANT: Stockinger, Eric

APPLICANT: Jaglo-Ottosen, Kirsten
APPLICANT: Gilmore, Sarah
APPLICANT: Zarka, Daniel
APPLICANT: Jiang, Cai-Zhong
TITLE OF INVENTION: Plant Having Altered Environmental Stress Tolerance
FILE REFERENCE: 19117.713 Seq List
CURRENT APPLICATION NUMBER: US/09/198.119C
CURRENT FILING DATE: 1998-11-23
PRIOR APPLICATION NUMBER: US 08/706,270
PRIOR FILING DATE: 1998-09-04
PRIOR APPLICATION NUMBER: US 09/018,233
PRIOR FILING DATE: 1998-02-03
PRIOR APPLICATION NUMBER: US 09/017,816
PRIOR FILING DATE: 1998-02-03
PRIOR APPLICATION NUMBER: US 09/018,235
PRIOR FILING DATE: 1998-02-03
PRIOR APPLICATION NUMBER: US 09/017,575
PRIOR FILING DATE: 1998-02-03
PRIOR APPLICATION NUMBER: US 09/018,227
PRIOR FILING DATE: 1998-02-03
PRIOR APPLICATION NUMBER: US 09/018,234
PRIOR FILING DATE: 1998-02-03
NUMBER OF SEQ ID NOS: 95
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 56
LENGTH: 1132
TYPE: DNA
ORGANISM: Brassica napus
FEATURE:
OTHER INFORMATION: bnCBF6 gene
US-09-198-119C-56

Query Match 1.1%; Score 18; DB 4; Length 1132;
Best Local Similarity 100.0%; Pred.No. 27;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1364 CCCCATTTGACAGATGGT 1381
DB 851 CCCCATTTGACAGATGGT 834

RESULT 8
US-09-249-180-5/c
Sequence 5, Application US/09249180
Patent No. 6268548
GENERAL INFORMATION:
APPLICANT: Elthon, Thomas E
APPLICANT: Lund, Adrian A
APPLICANT: Bhatramakki, Dinakar
APPLICANT: Rhoads, David M.
TITLE OF INVENTION: Isolation and Characterization of Heat Shock Protein
FILE REFERENCE: UNV52819
CURRENT APPLICATION NUMBER: US/09/249,180
CURRENT FILING DATE: 1999-02-12
EARLIER APPLICATION NUMBER: 60/076/014
EARLIER FILING DATE: 1998-02-26
NUMBER OF SEQ ID NOS: 7
SOFTWARE: Created in Patent In Ver. 2.0, Edited in WordPerfect 6.1
SEQ ID NO 5
LENGTH: 1163
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: CDS
LOCATION: (79)..(337)
FEATURE:
NAME/KEY: CDS
LOCATION: (447)..(970)
FEATURE:
NAME/KEY: exon
LOCATION: (1)..(337)
FEATURE:
NAME/KEY: exon

LOCATION: (447)..(1163)
FEATURE:
NAME/KEY: gene
LOCATION: (1)..(1163)
FEATURE:
NAME/KEY: intron
LOCATION: (329)..(463)
FEATURE:
NAME/KEY: misc_feature
LOCATION: (559)..(828)
OTHER INFORMATION: Heat Shock Domain
FEATURE:
NAME/KEY: mRNA
LOCATION: (1)..(337)
FEATURE:
NAME/KEY: mRNA
LOCATION: (447)..(1163)
FEATURE:
NAME/KEY: Poly A site
LOCATION: (1163)..(1163)
FEATURE:
NAME/KEY: source
LOCATION: (1)..(1163)
OTHER INFORMATION: Zea Mays L., Line B73
FEATURE:
NAME/KEY: transit peptide
LOCATION: (79)..(213)
FEATURE:
NAME/KEY: 5'UTR
LOCATION: (1)..(78)
FEATURE:
NAME/KEY: 3'UTR
LOCATION: (871)..(1163)
US-09-249-180-5

Query Match 1.1%; Score 18; DB 3; Length 1163;
Best Local Similarity 100.0%; Pred.No. 27;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1391 ATACTGATGAGATAGATA 1408
DB 1015 ATACTGATGAGATAGATA 998

RESULT 9
US-09-328-352-2704/c
Sequence 2704, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 2704
LENGTH: 2148
TYPE: DNA
ORGANISM: Acinetobacter baumannii
US-09-328-352-2704

Query Match 1.1%; Score 18; DB 4; Length 2148;
Best Local Similarity 100.0%; Pred.No. 27;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 363 ATGATGGTATGCTGATG 380
DB 1263 ATGATGGTATGCTGATG 1246

RESULT 10
US-08-545-528D-1

Sequence 1, Application US/08545528D
Patent No. 6537773
GENERAL INFORMATION:

APPLICANT: FRASER et al.
TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment
Patent No. 6537773

TITLE OF INVENTION: Thereof, and Uses Thereof

FILE REFERENCE: PB193P1

CURRENT APPLICATION NUMBER: US/08/545,528D

CURRENT FILING DATE: 1995-10-19

PRIOR APPLICATION NUMBER: US 08/488,018

PRIOR FILING DATE: 1995-06-07

PRIOR APPLICATION NUMBER: US 08/473,545

PRIOR FILING DATE: 1995-06-07

NUMBER OF SEQ ID NOS: 1

SOFTWARE: PatentIn version 3.1

SEQ ID NO 1

LENGTH: 580073

TYPE: DNA

ORGANISM: Mycoplasma genitalium

JS-08-545-528D-1

Query Match 1.1%; Score 18; DB 4; Length 580073;

Best Local Similarity 100.0%; Pred. No. 23;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 1332 ATTTGTGAAAGAGAAA 1349

|||||

b 355690 ATTTGTGAAAGAGAAA 355707

RESULT 11

JS-09-103-840A-2/c

Sequence 2, Application US/09103840A

Patent No. 6294328

GENERAL INFORMATION:

APPLICANT: FLEISCHMAN, Robert D.

APPLICANT: WHITE, Owen R.

APPLICANT: FRASER, Claire M.

APPLICANT: VENTER, John C.

TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

TITLE OF INVENTION: TUBERCULOSIS

FILE REFERENCE: 24366-20007.00

CURRENT APPLICATION NUMBER: US/09/103,840A

CURRENT FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 2

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 2

LENGTH: 4403765

TYPE: DNA

ORGANISM: Mycobacterium tuberculosis

FEATURE:

OTHER INFORMATION: CDC 1551

OTHER INFORMATION: "n" bases at various positions throughout the sequence

OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2

Query Match

Best Local Similarity 1.1%; Score 18; DB 3; Length 4403765;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1636 CGTCGTCATCTCCACCGT 1653

|||||

DB 4146706 CGTCGTCATCTCCACCGT 4146689

RESULT 12

US-09-103-840A-1/c

Sequence 1, Application US/09103840A

Patent No. 6294328

GENERAL INFORMATION:

APPLICANT: FLEISCHMAN, Robert D.

APPLICANT: WHITE, Owen R.

APPLICANT: FRASER, Claire M.

APPLICANT: VENTER, John C.

TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

TITLE OF INVENTION: TUBERCULOSIS

FILE REFERENCE: 24366-20007.00

CURRENT APPLICATION NUMBER: US/09/103,840A

CURRENT FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 2

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 1

LENGTH: 4411529

TYPE: DNA

ORGANISM: Mycobacterium tuberculosis

OTHER INFORMATION: H37Rv

US-09-103-840A-1

Query Match

Best Local Similarity 1.1%; Score 18; DB 3; Length 4411529;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1636 CGTCGTCATCTCCACCGT 1653

|||||

DB 4154513 CGTCGTCATCTCCACCGT 4154496

RESULT 13

US-08-120-827-71/c

Sequence 71, Application US/08120827

Patent No. 5525495

GENERAL INFORMATION:

APPLICANT: KEENE, JACK D.

APPLICANT: KING, PETER H.

APPLICANT: LEVINE, TODD

TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE

TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS

TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION

NUMBER OF SEQUENCES: 101

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

STREET: 1755 Jefferson Davis Highway, Fourth Floor

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/120,827

FILING DATE: 15-SEP-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Oblon, No. 5525495man F.

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 714-158-0 CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)413-3000

TELEFAX: (703)413-2220

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 71:

SEQUENCE CHARACTERISTICS:

LENGTH: 28 base pairs

TYPE: nucleic acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: RNA (genomic)

US-08-120-827-71

Query Match

Best Local Similarity 1.0%; Score 17; DB 1; Length 28;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1315 AACAAATAAAGAAAAC 1331
Db 28 AACAAATAAAGAAAAC 12

RESULT 14
US-08-478-675-71/c
; Sequence 71, Application US/08478675
; Patent No. 5773246
; GENERAL INFORMATION:
; APPLICANT: KEENE, JACK D.
; APPLICANT: KING, PETER H.
; APPLICANT: LEVINE, TODD
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE
; TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACIDS
; TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION
; NUMBER OF SEQUENCES: 101
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,675
; FILING DATE: 07-JUN-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/120,827
; FILING DATE: 15-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5773246man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 714-158-0 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)413-3000
; TELEFAX: (703)413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: RNA (genomic)
US-08-478-675-71

Query Match 1.0%; Score 17; DB 1; Length 28;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1315 AACAAATAAAGAAAAC 1331
Db 28 AACAAATAAAGAAAAC 12

RESULT 15
US-09-134-001C-2588/c
; Sequence 2588, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2588
; LENGTH: 210
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2588

Query Match 1.0%; Score 17; DB 4; Length 210;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 221 AAACAAGCATCATTAAA 237
Db 95 AAACAAGCATCATTAAA 79

Search completed: February 2, 2004, 07:59:13
Job time : 134.678 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 2, 2004, 03:00:32 ; Search time 1254.7 Seconds
(without alignments)
4871.226 Million cell updates/sec

Title: US-09-938-842A-3729

Perfect score: 1677

Sequence: 1 GGTAAAGCGTTTACTTATG.....TTTCTCTCAGTATATTGA 1677

Scoring table: OLIGO_NUC

GAPOP 60.0 , Gapext 60.0

Searched: 2434939 seqs, 1822278265 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4869878

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
- 14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|---|
| 1 | 1677 | 100.0 | 1677 | 10 | US-09-938-842A-3729 Sequence 3729, Ap |
| 2 | 1677 | 100.0 | 1677 | 12 | US-09-938-842A-3729 Sequence 3729, Ap |
| 3 | 21 | 1.3 | 330 | 13 | US-09-930-213-54 Sequence 54, Appl |
| 4 | 21 | 1.3 | 394 | 11 | US-09-803-719-707 Sequence 707, Appl |
| 5 | 21 | 1.3 | 407 | 11 | US-09-803-719-1633 Sequence 1633, Ap |
| 6 | 21 | 1.3 | 2000 | 13 | US-09-930-213-306 Sequence 306, Appl |
| C 7 | 21 | 1.3 | 108359 | 15 | US-10-191-807-3 Sequence 3, Appli |
| | 20 | 1.2 | 627 | 13 | US-10-027-632-54726 Sequence 54726, A |
| C 8 | 20 | 1.2 | 627 | 13 | US-10-027-632-294129 Sequence 294129, A |
| | 20 | 1.2 | 627 | 14 | US-10-027-632-54726 Sequence 54726, A |
| C 9 | 20 | 1.2 | 627 | 14 | US-10-027-632-294129 Sequence 294129, A |
| C 10 | 20 | 1.2 | 627 | 14 | US-10-027-632-294129 Sequence 294129, A |
| C 11 | 20 | 1.2 | 768 | 9 | US-09-815-242-9987 Sequence 9987, Ap |
| C 12 | 20 | 1.1 | 199 | 13 | US-10-029-386-14103 Sequence 14103, A |
| C 13 | 19 | 1.1 | 199 | 13 | US-10-029-386-14103 Sequence 14103, A |
| C 14 | 19 | 1.1 | 281 | 12 | US-10-242-535A-18166 Sequence 18166, A |
| C 15 | 19 | 1.1 | 377 | 10 | US-09-983-965-5678 Sequence 5678, Ap |

| | | | | | |
|------|----|-----|------|----|---|
| C 16 | 19 | 1.1 | 400 | 8 | US-08-781-986A-3898 Sequence 3898, Ap |
| C 17 | 19 | 1.1 | 547 | 13 | US-10-029-386-398 Sequence 398, App |
| C 18 | 19 | 1.1 | 588 | 13 | US-10-027-632-88862 Sequence 88862, A |
| C 19 | 19 | 1.1 | 588 | 13 | US-10-027-632-88863 Sequence 88863, A |
| C 20 | 19 | 1.1 | 588 | 14 | US-10-027-632-88862 Sequence 88862, A |
| C 21 | 19 | 1.1 | 588 | 14 | US-10-027-632-88863 Sequence 88863, A |
| C 22 | 19 | 1.1 | 598 | 13 | US-10-027-632-230239 Sequence 230239, A |
| C 23 | 19 | 1.1 | 598 | 13 | US-10-027-632-230240 Sequence 230240, A |
| C 24 | 19 | 1.1 | 598 | 14 | US-10-027-632-230239 Sequence 230239, A |
| C 25 | 19 | 1.1 | 598 | 14 | US-10-027-632-230240 Sequence 230240, A |
| C 26 | 19 | 1.1 | 652 | 13 | US-10-027-632-133067 Sequence 133067, A |
| C 27 | 19 | 1.1 | 652 | 14 | US-10-027-632-133067 Sequence 133067, A |
| C 28 | 19 | 1.1 | 672 | 13 | US-10-027-632-34038 Sequence 34038, A |
| C 29 | 19 | 1.1 | 672 | 14 | US-10-027-632-34038 Sequence 34038, A |
| C 30 | 19 | 1.1 | 903 | 9 | US-09-815-242-3975 Sequence 2975, Ap |
| C 31 | 19 | 1.1 | 909 | 9 | US-09-815-242-4338 Sequence 4338, Ap |
| C 32 | 19 | 1.1 | 921 | 9 | US-09-815-242-8202 Sequence 8202, Ap |
| C 33 | 19 | 1.1 | 990 | 13 | US-10-027-632-262153 Sequence 262153, A |
| C 34 | 19 | 1.1 | 990 | 13 | US-10-027-632-262154 Sequence 262154, A |
| C 35 | 19 | 1.1 | 990 | 14 | US-10-027-632-262153 Sequence 262153, A |
| C 36 | 19 | 1.1 | 990 | 14 | US-10-027-632-262154 Sequence 262154, A |
| C 37 | 19 | 1.1 | 1011 | 13 | US-10-027-632-122375 Sequence 122375, A |
| C 38 | 19 | 1.1 | 1011 | 14 | US-10-027-632-122375 Sequence 122375, A |
| C 39 | 19 | 1.1 | 1103 | 13 | US-10-027-632-30495 Sequence 30495, A |
| C 40 | 19 | 1.1 | 1103 | 13 | US-10-027-632-30496 Sequence 30496, A |
| C 41 | 19 | 1.1 | 1103 | 14 | US-10-027-632-30495 Sequence 30495, A |
| C 42 | 19 | 1.1 | 1103 | 14 | US-10-027-632-30496 Sequence 30496, A |
| C 43 | 19 | 1.1 | 2000 | 10 | US-09-938-842A-3833 Sequence 3833, Ap |
| C 44 | 19 | 1.1 | 2000 | 12 | US-09-938-842A-3833 Sequence 3833, Ap |
| C 45 | 19 | 1.1 | 2352 | 10 | US-09-938-842A-1592 Sequence 1592, Ap |

ALIGNMENTS

RESULT 1

US-09-938-842A-3729
; Sequence 3729, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kresps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME AND METHODS OF USE
; FILE REFERENCE: SCHIP1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 3729
; LENGTH: 1677
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-3729

| | | | | |
|-----------------------|--------------|---|------------|-------------|
| Query Match | 100.0% | Score 1677 | DB 10 | Length 1677 |
| Best Local Similarity | 100.0% | Pred. No. 0 | | |
| Matches 1677 | Conservative | 0 | Mismatches | 0 |
| | | 0 | Indels | 0 |
| | | 0 | Gaps | 0 |
| QY | 1 | GGTAAAGCGTTTACTTATGTTATATGCAAGGAGATATGCAATGTTGGAATGC | 60 | |
| DB | 1 | GGTAAAGCGTTTACTTATGTTATATGCAAGGAGATATGCAATGTTGGAATGC | 60 | |
| QY | 61 | TTTTTCAGATCATCAAAAGGTCCTACAGATTTCTAGGAAATGTTTCAGGCTTTGTGA | 120 | |
| DB | 61 | TTTTTCAGATCATCAAAAGGTCCTACAGATTTCTAGGAAATGTTTCAGGCTTTGTGA | 120 | |

[illegible]

| | | |
|------|--|------|
| 1201 | AAAGCTAGGTATTTTCATTTGGGAGTGTACTAGTAACCTAGTAACTAACCCAGAATGAGT | 1260 |
| 1261 | TTCTGATTTTGGATTTTGAAGCTTTTCTTAGGTTAAAAACAAGATATATTTACTAAACAAT | 1320 |
| 1261 | TTCTGATTTTGGATTTTGAAGCTTTTCTTAGGTTAAAAACAAGATATATTTACTAAACAAT | 1320 |
| 1321 | AAAAAGAAAAACATTTTGTGAAAGAAGAAATAAAGTTTACTGGACCCCAATTGTACAGATGG | 1380 |
| 1321 | AAAAAGAAAAACATTTTGTGAAAGAAGAAATAAAGTTTACTGGACCCCAATTGTACAGATGG | 1380 |
| 1381 | TCCCATATAATACCTGATAGAGATAGAGCAATGGAAGTGTATTTGTTTCACGTGGTACAA | 1440 |
| 1381 | TCCCATATAATACCTGATAGAGATAGAGCAATGGAAGTGTATTTGTTTCACGTGGTACAA | 1440 |
| 1441 | TCGGAAATGGTCTTTTAAAGCTCATCGAAACATCATCAGGACCGTTGATTTTTCCCGCATCAA | 1500 |
| 1441 | TCGGAAATGGTCTTTTAAAGCTCATCGAAACATCATCAGGACCGTTGATTTTTCCCGCATCAA | 1500 |
| 1501 | AAAGGGTTGAATACATTCTACATTGTTTTTCTGCTCCTATATATATATATCTGACGAGTC | 1560 |
| 1501 | AAAGGGTTGAATACATTCTACATTGTTTTTCTGCTCCTATATATATATATCTGACGAGTC | 1560 |
| 1561 | ACATTTAGTAATCTCCTTTGGACGTGTAACGGCGTTAAAAACGATCTTTTCCCATTTGTAATCC | 1620 |
| 1561 | ACATTTAGTAATCTCCTTTGGACGTGTAACGGCGTTAAAAACGATCTTTTCCCATTTGTAATCC | 1620 |
| 1621 | GTTTTTAAACAATCTGTCGTGCATCTCAACCGTTCGGTTTTTCTCAGCTATATTTTA | 1677 |
| 1621 | GTTTTTAAACAATCTGTCGTGCATCTCAACCGTTCGGTTTTTCTCAGCTATATTTTA | 1677 |

RESULT 3
US-09-930-213-54
Sequence 54, Application US/09930213
Publication No. US20030170625A1
GENERAL INFORMATION:
APPLICANT: ROSENTHAL, ANDRE
APPLICANT: HINZMANN, BERND
APPLICANT: SCHAFER, REINHARD
APPLICANT: ZUBER, JOHANNES
APPLICANT: TCHE-NITSE, OLEG
APPLICANT: GRIPS, MARTIN
APPLICANT: HELNEGEL, MARTIN
APPLICANT: SCHMITZ, ANNE-CHANTAL
APPLICANT: SERS, CHRISTINE
TITLE OF INVENTION: DETECTION OF DIFFERENTIAL GENE EXPRESSIONS
FILE REFERENCE: ALBRE-14
CURRENT APPLICATION NUMBER: US/09/930,213
CURRENT FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: DE 10004102.7
PRIOR FILING DATE: 2000-01-31
NUMBER OF SEQ ID NOS: 885
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 54
LENGTH: 330
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: modified_base
LOCATION: (4)
OTHER INFORMATION: a, t, c, g, other or unknown
FEATURE:
NAME/KEY: modified_base
LOCATION: (9)
OTHER INFORMATION: a, t, c, g, other or unknown
US-09-930-213-54

Query Match 1.3%; Score 21; DB 13; Length 330;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 416 TTATCCCTCTGTGTGACCC 436

Db 148 TTATCCCTCTGTGGTGACCC 168

RESULT 4

US-09-803-719-707
; Sequence 707, Application US/09803719
; Publication No. US20030044783A1
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Escobedo, Jaime
; APPLICANT: Innis, Michael A.
; APPLICANT: Garcia, Pablo Dominguez
; APPLICANT: Sudduth-Klinger, Julie
; APPLICANT: Reinhard, Christoph
; APPLICANT: Giese, Klaus
; APPLICANT: Randazzo, Filippo
; APPLICANT: Kennedy, Giulia C.
; APPLICANT: Pot, David
; APPLICANT: Kassam, Altaf
; APPLICANT: Lamson, George
; APPLICANT: Drmanac, Radoje
; APPLICANT: Crkvenjakov, Radomir
; APPLICANT: Dickson, Mark
; APPLICANT: Drmanac, Snezana
; APPLICANT: Labat, Ivan
; APPLICANT: Leshkowitz, Dena
; APPLICANT: Kita, David
; APPLICANT: Garcia, Veronica
; APPLICANT: Jones, Lee William
; APPLICANT: Stache-Crain, Birgit
; TITLE OF INVENTION: Human Genes and Gene Products
; FILE REFERENCE: 1624.002
; CURRENT APPLICATION NUMBER: US/09/803,719
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/188,609
; NUMBER OF SEQ ID NOS: 2396
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 707
; LENGTH: 394
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-803-719-707

Query Match 1.3%; Score 21; DB 11; Length 394;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 416 TTATCCCTCTGTGGTGACCC 436
Db 49 TTATCCCTCTGTGGTGACCC 69

RESULT 5

US-09-803-719-1633
; Sequence 1633, Application US/09803719
; Publication No. US20030044783A1
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Escobedo, Jaime
; APPLICANT: Innis, Michael A.
; APPLICANT: Garcia, Pablo Dominguez
; APPLICANT: Sudduth-Klinger, Julie
; APPLICANT: Reinhard, Christoph
; APPLICANT: Giese, Klaus
; APPLICANT: Randazzo, Filippo
; APPLICANT: Kennedy, Giulia C.
; APPLICANT: Pot, David
; APPLICANT: Kassam, Altaf
; APPLICANT: Lamson, George
; APPLICANT: Drmanac, Radoje
; APPLICANT: Crkvenjakov, Radomir
; APPLICANT: Dickson, Mark

; APPLICANT: Drmanac, Snezana
; APPLICANT: Labat, Ivan
; APPLICANT: Leshkowitz, Dena
; APPLICANT: Kita, David
; APPLICANT: Garcia, Veronica
; APPLICANT: Jones, Lee William
; APPLICANT: Stache-Crain, Birgit
; TITLE OF INVENTION: Human Genes and Gene Products
; FILE REFERENCE: 1624.002
; CURRENT APPLICATION NUMBER: US/09/803,719
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/188,609
; NUMBER OF SEQ ID NOS: 2396
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1633
; LENGTH: 407
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-803-719-1633

Query Match 1.3%; Score 21; DB 11; Length 407;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 416 TTATCCCTCTGTGGTGACCC 436
Db 195 TTATCCCTCTGTGGTGACCC 215

RESULT 6

US-09-930-213-306
; Sequence 306, Application US/09930213
; Publication No. US20030170625A1
; GENERAL INFORMATION:
; APPLICANT: ROSENTHAL, ANDRE
; APPLICANT: HINZMANN, BERND
; APPLICANT: SCHAFER, REINHARD
; APPLICANT: ZUBER, JOHANNES
; APPLICANT: TCHE-NITSE, OLEG
; APPLICANT: GRIPS, MARTIN
; APPLICANT: HELLMER, MARTIN
; APPLICANT: SCHMITZ, ANNE-CHANTAL
; APPLICANT: SERS, CHRISTINE
; TITLE OF INVENTION: DETECTION OF DIFFERENTIAL GENE EXPRESSIONS
; FILE REFERENCE: ALBRE-14
; CURRENT APPLICATION NUMBER: US/09/930,213
; CURRENT FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: DE 10004102.7
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 885
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 306
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-930-213-306

Query Match 1.3%; Score 21; DB 13; Length 2000;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 416 TTATCCCTCTGTGGTGACCC 436
Db 885 TTATCCCTCTGTGGTGACCC 905

RESULT 7

US-10-191-807-3/c
; Sequence 3, Application US/10191807
; Publication No. US20030068691A1
; GENERAL INFORMATION:
; APPLICANT: HU, Song et al.

;; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
;; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
;; TITLE OF INVENTION: US\$ THEREOF
;; FILE REFERENCE: CL001273-PROV
;; CURRENT APPLICATION NUMBER: US/10/191.807
;; CURRENT FILING DATE: 2002-07-10
;; NUMBER OF SEQ ID NOS: 5
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 3
;; LENGTH: 108359
;; TYPE: DNA
;; ORGANISM: Human
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: (1)-(108359)
;; OTHER INFORMATION: n = A,T,C or G

JS-10-191-807-3

Query Match 1.3%; Score 21; DB 15; Length 108359;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 240 TTAATGGTTTGAAGAATA 260
|||
b 75869 TTAATGGTTTGAAGAATA 75849

RESULT 8

JS-10-027-632-54726/c
; Sequence 54726, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.

;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;; TITLE OF INVENTION: Polymorphisms in the Human Genome
;; FILE REFERENCE: 108827.129
;; CURRENT APPLICATION NUMBER: US/10/027,632
;; CURRENT FILING DATE: 2002-04-30
;; PRIOR APPLICATION NUMBER: US 60/218,006
;; PRIOR FILING DATE: 2000-07-12
;; PRIOR APPLICATION NUMBER: US 60/198,676
;; PRIOR FILING DATE: 2000-04-20
;; PRIOR APPLICATION NUMBER: US 60/193,483
;; PRIOR FILING DATE: 1999-09-28
;; PRIOR APPLICATION NUMBER: US 60/156,358
;; PRIOR FILING DATE: 1999-08-09
;; NUMBER OF SEQ ID NOS: 325720
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 54726

;; LENGTH: 627
;; TYPE: DNA
;; ORGANISM: Human
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: (1)-(627)
;; OTHER INFORMATION: n = A,T,C or G

JS-10-027-632-54726

Query Match 1.2%; Score 20; DB 13; Length 627;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 1134 AAAAATTATTGAAATCTTT 1153
|||
b 508 AAAAATTATTGAAATCTTT 489

RESULT 9

US-10-027-632-294129/c
; Sequence 294129, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.

;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;; TITLE OF INVENTION: Polymorphisms in the Human Genome
;; FILE REFERENCE: 108827.129
;; CURRENT APPLICATION NUMBER: US/10/027,632
;; CURRENT FILING DATE: 2002-04-30
;; PRIOR APPLICATION NUMBER: US 60/218,006
;; PRIOR FILING DATE: 2000-07-12
;; PRIOR APPLICATION NUMBER: US 60/198,676
;; PRIOR FILING DATE: 2000-04-20
;; PRIOR APPLICATION NUMBER: US 60/193,483
;; PRIOR FILING DATE: 2000-03-29
;; PRIOR APPLICATION NUMBER: US 60/185,218
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: US 60/167,363
;; PRIOR FILING DATE: 1999-11-23
;; PRIOR APPLICATION NUMBER: US 60/156,358
;; PRIOR FILING DATE: 1999-09-28
;; PRIOR APPLICATION NUMBER: US 60/146,002
;; PRIOR FILING DATE: 1999-08-09
;; NUMBER OF SEQ ID NOS: 325720
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 294129

;; LENGTH: 627
;; TYPE: DNA
;; ORGANISM: Human
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: (1)-(627)
;; OTHER INFORMATION: n = A,T,C or G

US-10-027-632-294129

Query Match 1.2%; Score 20; DB 13; Length 627;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 1134 AAAAATTATTGAAATCTTT 1153
|||
b 508 AAAAATTATTGAAATCTTT 489

RESULT 10

US-10-027-632-54726/c
; Sequence 54726, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.

;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;; TITLE OF INVENTION: Polymorphisms in the Human Genome
;; FILE REFERENCE: 108827.129
;; CURRENT APPLICATION NUMBER: US/10/027,632
;; CURRENT FILING DATE: 2002-04-30
;; PRIOR APPLICATION NUMBER: US 60/218,006
;; PRIOR FILING DATE: 2000-07-12
;; PRIOR APPLICATION NUMBER: US 60/198,676
;; PRIOR FILING DATE: 2000-04-20
;; PRIOR APPLICATION NUMBER: US 60/193,483
;; PRIOR FILING DATE: 2000-03-29
;; PRIOR APPLICATION NUMBER: US 60/185,218
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: US 60/167,363
;; PRIOR FILING DATE: 1999-11-23
;; PRIOR APPLICATION NUMBER: US 60/156,358
;; PRIOR FILING DATE: 1999-09-28
;; PRIOR APPLICATION NUMBER: US 60/146,002
;; PRIOR FILING DATE: 1999-08-09
;; NUMBER OF SEQ ID NOS: 325720
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 54726

;; LENGTH: 627
;; TYPE: DNA
;; ORGANISM: Human
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: (1)-(627)
;; OTHER INFORMATION: n = A,T,C or G

LENGTH: 627
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(627)
OTHER INFORMATION: n = A,T,C or G
US-10-027-632-54726

Query Match
Best Local Similarity 100.0%; Pred. No. 25; Length 627;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 AAAAATTATTGAAATCTTT 1153
DB 508 AAAAATTATTGAAATCTTT 489

RESULT 11
US-10-027-632-294129/c
Sequence 294129, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
POLYMORPHISMS IN THE HUMAN GENOME
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 294129
LENGTH: 627
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(627)
OTHER INFORMATION: n = A,T,C or G
US-10-027-632-294129

Query Match
Best Local Similarity 100.0%; Pred. No. 25; Length 627;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1134 AAAAATTATTGAAATCTTT 1153
DB 508 AAAAATTATTGAAATCTTT 489

RESULT 12
US-09-815-242-9987/c
Sequence 9987, Application US/09815242
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
POLYMORPHISMS IN THE HUMAN GENOME
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 294129
LENGTH: 627
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(627)
OTHER INFORMATION: n = A,T,C or G
US-10-027-632-294129

APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
THE HUMAN GENOME
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9987
LENGTH: 768
TYPE: DNA
ORGANISM: Salmonella typhi
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(768)
US-09-815-242-9987

Query Match
Best Local Similarity 100.0%; Pred. No. 26; Length 768;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1249 ACCAGAATGAGTTTCGATT 1268
DB 332 ACCAGAATGAGTTTCGATT 313

RESULT 13
US-10-029-386-14103/c
Sequence 14103, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
IDENTIFYING GENES
FILE REFERENCE: AEONICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 14103
LENGTH: 199
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL135926.3
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2
OTHER INFORMATION: EST HUMAN HIT: AU152688.1, EVALUATE 1.00e-06
OTHER INFORMATION: SWISSPROT HIT: Q09332, EVALUATE 7.30e+00
OTHER INFORMATION: NT HIT: U49016.1, EVALUATE 2.00e-10
US-10-029-386-14103

US-10-029-386-14103/c
Sequence 14103, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
IDENTIFYING GENES
FILE REFERENCE: AEONICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 14103
LENGTH: 199
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL135926.3
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2
OTHER INFORMATION: EST HUMAN HIT: AU152688.1, EVALUATE 1.00e-06
OTHER INFORMATION: SWISSPROT HIT: Q09332, EVALUATE 7.30e+00
OTHER INFORMATION: NT HIT: U49016.1, EVALUATE 2.00e-10
US-10-029-386-14103

```
Query Match      1.1%; Score 19; DB 13; Length 199;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1259 GTTCTCGATTTCGATTTT 1277
Db 63 GTTCTCGATTTCGATTTT 45

RESULT 14
US-10-242-535A-18166/c
; Sequence 18166, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18166
; LENGTH: 281
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (31)..(31)
; OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-18166

Query Match      1.1%; Score 19; DB 12; Length 281;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 702 TGAGCCCTTCTTTCTGG 720
Db 64 TGAGCCCTTCTTTCTGG 46

RESULT 15
US-09-983-965-5678
; Sequence 5678, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 5678
; LENGTH: 377
; TYPE: DNA
; ORGANISM: Bos taurus
```

```
; FEATURE:
; OTHER INFORMATION: Clone ID: 55-LIB34-044-Q1-E1-F4
US-09-983-965-5678
```

```
Query Match      1.1%; Score 19; DB 10; Length 377;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1516 ATTCTCACTTGTTCCTG 1534
Db 158 ATTCTCACTTGTTCCTG 176
```

```
Search completed: February 2, 2004, 08:33:18
Job time : 1258.7 secs
```